Online ISSN : 0975-4172
Print ISSN : 0975-4350

Global Journal of Computer Science and Technology : C

Software and Data Engineering

DISCOVERING THOUGHTS AND INVENTING FUTURE

HIGHLIGHTS

- Encapsulation of Soft Computing
- Optimization of Job Scheduling
- Algorithm in Data Mining
- Artificial System for Prediction

Datacentre

Volume 12 | Issue 15 | Version 1.0 ENG

Global Journal of Computer Science and Technology: C
Software & Data Engineering

Global Journal of Computer Science and Technology: C
Software & Data Engineering

Volume 12 Issue 15 (Ver. 1.0)

Open Association of Research Society

Incorporation No.: 0423089
License No.: 42125/022010/1186
Registration No.: 430374
Import-Export Code: 1109007027
Employer Identification Number (EIN):
USA Tax ID: 98-0673427

Global Journals Inc.

(A Delaware USA Incorporation with "Good Standing"; Reg. Number: 0423089)

Sponsors:Global Association of Research

Open Scientific Standards

Publisher's Headquarters office

Global Journals Inc., Headquarters Corporate Office, Cambridge Office Center, II Canal Park, Floor No. 5th, ***Cambridge (Massachusetts)***, Pin: MA 02141 United States

USA Toll Free: +001-888-839-7392

USA Toll Free Fax: +001-888-839-7392

Offset Typesetting

Global Association of Research, Marsh Road, Rainham, Essex, London RM13 8EU United Kingdom.

Packaging & Continental Dispatching

Global Journals, India

Find a correspondence nodal officer near you

To find nodal officer of your country, please email us at *local@globaljournals.org*

eContacts

Press Inquiries: *press@globaljournals.org*
Investor Inquiries: *investers@globaljournals.org*
Technical Support: *technology@globaljournals.org*
Media & Releases: *media@globaljournals.org*

Pricing (Including by Air Parcel Charges):

For Authors:
22 USD (B/W) & 50 USD (Color)
Yearly Subscription (Personal & Institutional):
200 USD (B/W) & 250 USD (Color)

Editorial Board Members (HON.)

Dr. Bart Lambrecht
Director of Research in Accounting and FinanceProfessor of Finance
Lancaster University Management School
BA (Antwerp); MPhil, MA, PhD (Cambridge)

Dr. Carlos García Pont
Associate Professor of Marketing
IESE Business School, University of Navarra
Doctor of Philosophy (Management), Massachusetts Institute of Technology (MIT)
Master in Business Administration, IESE, University of Navarra
Degree in Industrial Engineering, Universitat Politècnica de Catalunya

Dr. Fotini Labropulu
Mathematics - Luther College
University of ReginaPh.D., M.Sc. in Mathematics
B.A. (Honors) in Mathematics
University of Windso

Dr. Lynn Lim
Reader in Business and Marketing
Roehampton University, London
BCom, PGDip, MBA (Distinction), PhD, FHEA

Dr. Mihaly Mezei
ASSOCIATE PROFESSOR
Department of Structural and Chemical Biology, Mount Sinai School of Medical Center
Ph.D., Etvs Lornd University
Postdoctoral Training,
New York University

Dr. Söhnke M. Bartram
Department of Accounting and FinanceLancaster University Management SchoolPh.D. (WHU Koblenz)
MBA/BBA (University of Saarbrücken)

Dr. Miguel Angel Ariño
Professor of Decision Sciences
IESE Business School
Barcelona, Spain (Universidad de Navarra)
CEIBS (China Europe International Business School).
Beijing, Shanghai and Shenzhen
Ph.D. in Mathematics
University of Barcelona
BA in Mathematics (Licenciatura)
University of Barcelona

Philip G. Moscoso
Technology and Operations Management
IESE Business School, University of Navarra
Ph.D in Industrial Engineering and Management, ETH Zurich
M.Sc. in Chemical Engineering, ETH Zurich

Dr. Sanjay Dixit, M.D.
Director, EP Laboratories, Philadelphia VA Medical Center
Cardiovascular Medicine - Cardiac Arrhythmia
Univ of Penn School of Medicine

Dr. Han-Xiang Deng
MD., Ph.D
Associate Professor and Research
Department Division of Neuromuscular Medicine
Davee Department of Neurology and Clinical NeuroscienceNorthwestern University
Feinberg School of Medicine

Dr. Pina C. Sanelli
Associate Professor of Public Health
Weill Cornell Medical College
Associate Attending Radiologist
NewYork-Presbyterian Hospital
MRI, MRA, CT, and CTA
Neuroradiology and Diagnostic Radiology
M.D., State University of New York at Buffalo,School of Medicine and Biomedical Sciences

Dr. Roberto Sanchez
Associate Professor
Department of Structural and Chemical Biology
Mount Sinai School of Medicine
Ph.D., The Rockefeller University

Dr. Wen-Yih Sun
Professor of Earth and Atmospheric SciencesPurdue University Director
National Center for Typhoon and Flooding Research, Taiwan
University Chair Professor
Department of Atmospheric Sciences, National Central University, Chung-Li, TaiwanUniversity Chair Professor
Institute of Environmental Engineering, National Chiao Tung University, Hsin-chu, Taiwan.Ph.D., MS The University of Chicago, Geophysical Sciences
BS National Taiwan University, Atmospheric Sciences
Associate Professor of Radiology

Dr. Michael R. Rudnick
M.D., FACP
Associate Professor of Medicine
Chief, Renal Electrolyte and Hypertension Division (PMC)
Penn Medicine, University of Pennsylvania
Presbyterian Medical Center, Philadelphia
Nephrology and Internal Medicine
Certified by the American Board of Internal Medicine

Dr. Bassey Benjamin Esu
B.Sc. Marketing; MBA Marketing; Ph.D Marketing
Lecturer, Department of Marketing, University of Calabar
Tourism Consultant, Cross River State Tourism Development Department
Co-ordinator , Sustainable Tourism Initiative, Calabar, Nigeria

Dr. Aziz M. Barbar, Ph.D.
IEEE Senior Member
Chairperson, Department of Computer Science
AUST - American University of Science & Technology
Alfred Naccash Avenue – Ashrafieh

Contents of the Volume

Global Journal of Computer Science and Technology
Software & Data Engineering
Volume 12 Issue 15 Version 1.0 Year 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Information System Development and Use Practices in Khyber Pakhtoon Khwa (K.P.K) Pakistan (An Empirical Study of the Demographics Impacts)

By Dr. Ghulam Muhammad Kundi & Dr. Allah Nawaz

Gomal University Dera Ismail Khan, K.P.K, Pakistan

Abstract - There is no doubt in the reality that Information Technology (IT) is revolutionizing organizations on unprecedented proportions thereby stimulating others to adopt it but despite this fact research indicates that a large number of information system development projects are failing to achieve their objectives in toto. Thus, there are partial and total failure stories of IT projects. Furthermore, information system (IS) failures are common to all types of organizations: public, private or small, medium and large irrespective of operating in a developed or developing country.

The research on IS failure frequently cites non-technical issues as the most decisive factors in the success or failure of any IT project. That is, IT can do miracles but all this requires 'adequate management of the 'demographics of an IT project.' Non-technical critical success and failure factors are catching wider attention during the last decades among the IS research community.

One can understand that technology can be imported but not the demographic of the organization thus, non-technical issues are 'local in nature, structure and intensity,' which definitely need local studies of ISD and use practices so as to dig-out 'customized ISD and use process. This research is an effort in the same line of thinking.

Keywords : information technology (IT), information system (IS), information system development (ISD), perception about IT, ISD approaches & methodologies, project management, ISDLC, success or failure, good and bad experiences, KPK pakistan.

GJCST-C Classification : H.4.0

Strictly as per the compliance and regulations of:

Information System Development and Use Practices in Khyber Pakhtoon Khwa (K.P.K) Pakistan

(*An Empirical Study of the Demographics Impacts*)

Dr. Ghulam Muhammad Kundi [α] & Dr. Allah Nawaz [σ]

Abstract - There is no doubt in the reality that Information Technology (IT) is revolutionizing organizations on unprecedented proportions thereby stimulating others to adopt it but despite this fact research indicates that a large number of information system development projects are failing to achieve their objectives in toto. Thus, there are partial and total failure stories of IT projects. Furthermore, information system (IS) failures are common to all types of organizations: public, private or small, medium and large irrespective of operating in a developed or developing country.

The research on IS failure frequently cites non-technical issues as the most decisive factors in the success or failure of any IT project. That is, IT can do miracles but all this requires 'adequate management of the 'demographics of an IT project.' Non-technical critical success and failure factors are catching wider attention during the last decades among the IS research community.

One can understand that technology can be imported but not the demographic of the organization thus, non-technical issues are 'local in nature, structure and intensity,' which definitely need local studies of ISD and use practices so as to dig-out 'customized ISD and use process. This research is an effort in the same line of thinking.

Keywords : information technology (IT), information system (IS), information system development (ISD), perception about IT, ISD approaches & methodologies, project management, ISDLC, success or failure, good and bad experiences, KPK pakistan.

I. Introduction

Developing a computer-based information system (CBIS) is not simply the purchase and installation of hardware and software (Rockart, et al., 1996; Smith, 1998; Walsham, 1993, 2000; Turban et al., 2004). It rather goes beyond it to the problems of people, organization and the context (Avgerou and Cornford, 1998); Segars and Grover, 1996); Dann et al., 1998). Research findings assert that an IS development (ISD) is a 'social process' (Lyytinen, 1987; Checkland, 1991; Walsham, 1993) thereby considering all the human, organizational, contextual and technological issues as in case of any organizational project.

Author α σ : Assistant Professor Department of Public Administration Gomal University Dera Ismail Khan, K.P.K, Pakistan.
E-mails : kundi@gu.edu.pk, profallahnawaz@yahoo.com

Human challenges include difference of perceptions about IT among the developers and users due to several gaps of education, communication, culture, motivation and satisfaction (see for example, Argyris, 1971; Kaasboll, 1997; Dann et al., 1998; Glass, 1998). The Nature (public or private), policies and procedures, the IT maturity, power structures etc., make up some of the issues emanating from the organization itself (see for example, Land et al., 1992; Ennals, 1995; Segars and Grover, 1996). Environment or context is significant since its change altogether changes requirements for the success/failure of an IT project (Flowers, 1997). Herzberg's two factors theory suggests that job-satisfiers relate to the job-contents while job-dissatisfiers emerge from the job-context (Luthans, 1995:149). Technology is not widely quoted as big deal but IT professionals are frequently cited as the toughest challenge in an IT project due To their intellectual distance from the nature and requirements of an organization (Argyris, 1971; Segars and Grover, 1996).

All of these challenges crop-up during different stages of an ISD life cycle (ISDLC). A global format for this cycle is: IS planning, requirements capture and analysis, design, implementation, use and maintenance and up-gradation (Avison and Wood-Harper, 1990). The intensity of issues vary from one stage to another for example, communication gap between developers and users at planning level is minor issue as compared to the same at requirements capture, training and use levels (Kaasboll, 1997). Likewise, organizational factors are less threatening during the initial phases of ISDLC but once new system is in action 'IT-business-alignment' (Burn, 1996; Poulymenko and Holemes, 1997) emerges as a big issue, which is widely reported as the major cause of many IS failure (ISF) cases around the world (see for example, Ewusi-Mensah and Przasnyski, 1991, 1994, 1995; Ennals, 1995; Glass, 1998).

In nutshell, IS community is unanimous on admitting that it is not technology-related issues rather human, organizational and contextual variables which make or break the future of an IT project (see for example, Avison and r-Harper, 1990; Poulymenakou and Holmes, 1996). Furthermore, all of these factors are

purely local in nature requiring customized-research projects to unearth indigenous footage of the impacts from these variables on the development trajectory of an IT projects.

The objectives of this study were to o unearth the ISD and use practices in KPK, Pakistan and local versions of challenges to the ISDLC from human, organizational, contextual and technology factors besides management concerns in the domestic IT-projects and to build-up a customized set of guidelines for handling an IT-project's development-trajectory successfully in education and health sectors of the economy.

This is the first project of its kind in KPK, Pakistan that unearthed purely 'localized and customized' problems and solution models for IT-projects. Likewise, the study will be contributive both in improving ISD and use practices as well as help in minimizing the chances of IS failure.

Since IT is indispensable to organizations but research warns that inadequate management of IT-projects result either into partial failure or total termination of the efforts. The question of this research therefore, was 'How far local management is succeeding in identifying and handling challenges to the ISD and use process in the indigenous context of KPK Pakistan?

II. Review of the Relevant Literature

The literature on IS development and use process is scattered across the organization, management, information-systems and computer studies. Researchers have identified critical success and failure factors (variables) about different aspects and stages in IT projects (see for example, Ennals, 1995; Beynon-Davies, 1995; Beynon-Davies and Lloyd-Williams, 1999). There is substantial evidence on the role of organizational, human and contextual factors in the whole process of infusing IT into organizational structure and culture (Walsham, 1993:25).

a) Demographics of ISD process

ISD is a social process therefore, it is certainty affected by all the surrounding factors. Organizational size and structure, policies, management style, methods and procedures, rules and regulations have to be taken into account at every step in the ISD and use process (Segars and Grover, 1996; Smith, 1998). Likewise, a fear-based organizational culture (Poulymenakou and Holmes, 1996) hinders a transparent IT project since people hesitate to admit mistakes and failures (Beynon-Davies, 1995; Warne, 1997; Beynon-Davies and Lloyd-Williams, 1999). An information system is designed, created, operated and used by humans thus, humans reflect in every move and dimension of the ISD and use trajectory (Sauer, 1993). Although technology (hardware, software and professionals) is neither an end nor all in the story of computerizing an organization, however, their availability and usability may trigger many questions. It is however, widely documented that IS developers (professionals) can create problems if developer-user gaps are not addressed early (Kaasboll, 1997).

b) Perceptions about IT

Life is what one believes in so perceptions of technology have bearing upon how they are used (see for example, Brooke, 1995; Collins and Bicknell, 1997). The perceptions of rich and poor nations have shifted away from economic milestones to knowledge yardsticks. Now information-rich and information-poor are the criteria to determine power of the nations. So where does a nation perceives itself on the continuum of digital-divide, reflects the use-level of IT in that country. In the organizational context, there are some kinds of 'silver-bullet' and 'leading-edge' syndromes (Ennals, 1995; Glass, 1998)' about IT expressing the belief that IT is a panacea for all management ills, while others disbelieve in any miraculous contributions of this technology (Baskerville and Smithson, 1995). Technocrats like accountants, engineers and scientists view IT as a commodity but mangers vision it, as a differentiator for the business.

c) Approaches and Methodologies

Several approaches have been theorized, exercised and reduced into black-n-white for computerization efforts (Hirschheim and Klein, 1989; Wynekoop and Russo 1995; Avison and Fitzgerald, 1995; Avison and Shah, 1997). They are grossly categorized into hard and soft approaches. Some researchers, particularly those hailing from computer science, suggest highly structured and scientifically managed approaches assuming that an IT project is a technical venture (Fitzgerald, 1996). Business managers however, perceive it as a business-project therefore prefer soft approaches so that the social nature of the development trajectory could be entertained (Walsham, 2000). These extremes have been compromised by the advocates of 'socio-technical' approaches, which assert that both technical and social management skills are required to handle IT-related efforts successfully.

Under hard and soft approaches, structured and unstructured ISD methodologies have been developed respectively. SSADM and STRAIDS (Weaver, 1993; DeMarco, 1979; Fitzgerald, 1996) are the structured examples while ETHICS (Mumford and Weir, 1979) and MultiView represent the soft methodologies. SSADM is most sophisticated and popular option. It is the official methodology for the public computerization projects in many countries, such as UK. Structured methodologies create technical and scientific behavior in the developers by offering techniques like highly structured DFDs, and CASE tools. Soft methodologies, on the other hand, give parallel importance to the

demographies like organization, humans and context. For example, ETHICS stands for effective technical and human implementation of computer systems. MultiView demands multi-view perception and treatment of computerization projects.

d) Project Management

An ISD and use process needs to be managed adequately otherwise leading-edge technology and huge budgets may gather dust. It is said that this adequacy is possible if it is recognized that "project is less a matter of understanding constraints and more a function of personal skills (Elton and Justin, 1998). Researchers have unearthed several IT-project management strategies. It is now squarely admitted that an IT project is like any other business project (Smith, 1998) therefore, all technical, organizational, human and contextual dimensions have to be brought on the table for visualizing a holistic view of the project.

e) ISD Life Cycle

An ISD process never ends since it demands constant upgrading thus, a cycle continues forever in the form of recursive stages (Avison and Fitzgerald, 1995; Avison and Shah, 1997; Turban et al, 2004:235). Several models are given in the literature to postulate a standard set of stages for an IT project. There are linear, waterfall and spiral models of an ISDLC. User-participation have widely been researched and identified as the critical factor to IT-project development and ultimately system's use (see for example, Mumford, and Henshall, 1979; Mumford, 1997). New CBIS changes the power structures therefore; losers and winners are created where losers naturally resist changing (see for example, Avison and Wood-Harper, 1990). An ISD has to be protected from the 'political maneuvering' or power struggle during the whole cycle of ISD otherwise, there is ample evidence on many IS failures, which were politically devastated (see for example, Markus, 1981, 1983; Drummond, 1995; McGrath, 1997).

f) Success or Failure (Good and Bad Experiences)

Literature is filled with stories of IT projects but unfortunately most episodes are about the failure because successes have little for research therefore reported occasionally (Glass, 1998). Failures are the repositories of the research questions for problem-solutions and improvement (Ewusi-Mensah and Przasnyski, 1991, 1992, 1994; McGrath, 1997). It has been found that the risk of IS failure is equal to all the small, medium and large enterprises in the developed and developing worlds and operating either in public or private sectors. IS failure have been extensively researched with the findings that there can be correspondence, process, system and expectation failures or project abandonment and terminations (Nawaz et al, 2007; Lyytinen and Hirschheim, 1987; Sauer, 1993; Ewusi-Mensah and Przasnyski, 1991, 1994, 1995). Whatever the name and nature of failure, there is broader agreement on two things: a. the same mistakes are committed in every IS failure case (Collins and Bicknell, 1997; Glass, 1998; Sauer, 1999) and b. the social, organizational, political and human factors outstrip the technical problems (Avgerou and Cornford, 1998).

III. Research Design

a) Survey Approach

Survey research is excessively used in information systems research (see for example, Galliers, 1992; Ewusi-Mensah Przasjyski, 1995; Olsen, 1997:23) as well as social research (Babbie, 256). Survey is preferred because in contrast to other approaches like, experiment, archival analysis and case studies, researcher can find answers to all five questions (who, what, where, how many, and how much) of the study (Yin, 1994:6) and thereby develop a comprehensive view of the problem. Furthermore, information systems as a field embodies a mixture of scientific, technical, organizational, societal and psychological aspects therefore, it is a multi-perspective discipline, which require 'pluralism of research methods' (Wood-Harper, 1985:169-191).

b) Population and Sampling

For the study of IS development and use practices in KPK Pakistan, Education and Health sectors were chosen for study on the pretext that 'both sectors have public and private organizations. Furthermore, Peshawar and DIKhan were sleeted for study on the ground that both the cities are totally different with reference to demographic variables. For example, Peshawar is a city of more than five million people while DIKhan has only 0.8m population. Likewise cities are extremely different in their organizational size and number, technological opportunities and applications and educational environment.'

Table 3.1 : Population & Samples from DIKhan, & Peshawar

	Sector	Org: Type	DIK	Pesh:	S-tot	Sample-Size		
						DIK	Pesh:	Tot
1	Health	Public	360	617	977	25	37	62
		Private	210	458	668	16	26	42
	S-tot		570	1075				104
2	Education	Public	625	720	1345	35	52	87
		Private	275	480	755	16	33	49
	S-tot		900	1200				136
3	Consultants		247	380	627	13	24	37
		Total	1717	2655	4372	105	172	277

Table 3.2 : Stratified Samples (Area-by-Sector Samples)

Area-Wise Sectors	Population (Strata)	Standard Deviations	Sample Sizes
Public Sector Health DIK	360	0.8	25
Public Sector Health Peshawar	617	0.7	37
Private Sector Health DIK	210	0.89	16
Private Sector Health Peshawar	458	0.66	26
Public Sector Education DIK	625	0.66	35
Public Sector Education Peshawar	720	0.87	52
Private Sector Education DIK	275	0.69	16
Private Sector Education Peshawar	480	0.8	33
Consultants DIK	247	0.6	13
Consultants Peshawar	380	0.78	24
TOTAL	4372		277

Table 3.3 : Sample Selection Procedures

Sample (FINITE population)

Pilot Study Statistics

Standard Deviation (□)	0.72
Standard Error (E)	0.082
Z value at 95% Confidence	1.96
Sample Population N	4372
Target Population	INFINITE
Sample Size	277

Formula

$$n = [\sigma^2/((z^2/E^2)+(\sigma^2/N))]$$

Stratified Samples

Pilot Study Statistics

	N	SD	N
Health (public) DIK	360	0.8	24
Health (public) Peshawar	617	0.7	37
Health (private) DIK	210	0.89	16
Health (private) Peshawar	458	0.66	26
Education (public) DIK	625	0.66	35
Education (public) Peshawar	720	0.87	53
Education (private) DIK	275	0.69	16
Education (private) Peshawar	480	0.8	33
Consultants DIK	247	0.6	13
Consultants Peshawar	380	0.78	25
N =	4372	n =	277

Formula

$$n_a = [(nN_a\sigma_a)/((N_a\sigma_a)+(N_b\sigma_b)+(N_c\sigma_c))]$$

c) Data Collection and Analysis

i. *Data Collection Methods*

Given the social-cum technical and global-cum-local nature of the topic, data was collected from all the possible sources to squarely cover all the related dimensions so that a comprehensive view of both the problem and solution could be envisaged.

1. *Literature Survey:* After preliminary literature survey for pilot study, the same was continued in the main research for two purposes: a. optimizing the selected variables and b. data on the topic.

2. *Self-administered Questionnaire:* It was the main inflow of primary data through a sophisticated and standardized set of questions arranged in a well-structured format. The instrument was successfully used in the pilot study. The same was applied in the main study.
3. *Follow-up Interviews:* Questionnaire covered the main variables; however, follow-up interviews were conducted for: a. collecting data that was missing in the questionnaire and b. gather data, which could not be captured through the questionnaire.

ii. *Data Analysis Tools*

Specific data analysis tools were used to carve-out meaning from the collected data. Tabulation was the top tool for 'data-reduction' as well as presentation of the findings. The tools used for analysis of the data in the study are given below:

1. *Descriptive Tools:* Besides textual analysis of secondary data, statistical descriptive-tools were used to explore and present: a. Respondents' profile (demographies) and b. Description of all the research-variables.
2. *Inferential Tools:* Correlation and regression analysis and significance tests were used to 'derive' meaning from data.

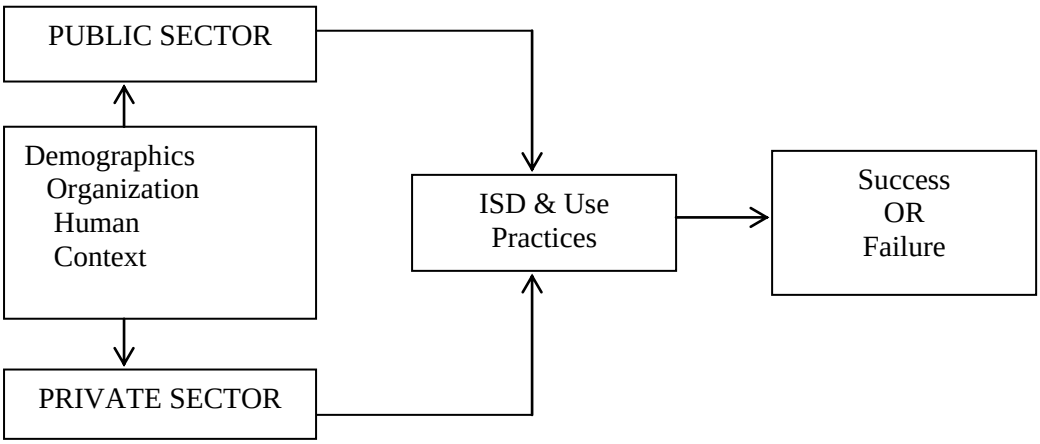
d) Operationalization of the Concepts

Table 3.4 : Operationalized Variables

	Variables	Attributes	Code
1	Human Factors	Perceptions about digital divide, silver-bullet syndrome, usability, commodity vs. differentiator, leading-edge syndrome, Organizational motivation techniques for IT, Perceptual gaps between developers and users.	HF
2	Organization Factors	Nature (public/private), Size, Structure, Objectives, and culture of the organization. IT maturity (experience with ISD and use) The mechanism for developer-user interaction Political/power struggles.	OF
3	Technological Factors	Hardware, Software and IT professionals. Availability, expenses, usability and possibility to upgrade the above items. Developers' organizational knowledge	TF
4	ISD	Government and Institutional IT Policies; User Needs Analysis; User Participation, Training; Implementation; Maintenance; and Evaluation in ISD, ISD approaches, Methodologies, Project management, User participation, developer-user communication, user training, Management of the resistance to change.	ISD
5	IS Use	Perceived Ease of Use (PEU); Perceived Usefulness (PU); Volume of Use; Experience with IT; User-developer- communication	USE
6	Perceptions	IT: the Problem-Solver; Digital Divide; and Socio-economic Impacts of IT.	PRC
7	Problems	Problems of IT Projects Development, Use and User-Satisfaction	PRB
8	Satisfaction	User-Satisfaction IT Projects Development and Use Practices.	STF
9	Opportunities	Opportunities for IT Project Success in K.P.K.	OPR
10	Success/ Failure	Definition of success/failure, Degree of success and failure, Ratio of success and failure, Critical success and failure factors, Escalation in IT projects.	SFF

e) Theoretical Framework

Figure 3.1 : Schematic Diagram of the Research-Model



f) Research Hypothesis

A set of hypothesis was developed on the basis of relationships postulated in the theoretical framework. Table 3.5 provides the detail.

Table 3.5 : List of Working Hypothesis

	Hypothesis	*Statistical Tools Applied*
1	The Public organizations are under-using IT potentials in comparison to private sector.	t-test
2	Escalation (time-delays, cost-overruns, compromise on lesser objectives) of IT projects is more common in public organizations than in private enterprises.	t-test
3	IT-people overestimate while non-IT workers underestimate the role of IT in the organizations.	t-test
4	Public sector is less optimistic about the role of IT than private sector.	t-test
5	Professors, doctors and consultants view IT differently.	ANOVA
6	Experience of non-IT workforce is negatively correlated with perceptions about IT.	Correlation analysis
7	Higher the perceptions about IT, greater are the chances/perceptions of success in IT projects	Simple Regression
8	The organizational, human, contextual and technological factors collectively determine the variation in the success/failure of an IT-project.	Multiple Regression

g) Pilot Study

All the above constructs and methods were used in the pilot study with the objectives of:

- Testing the research tools (particularly constructs). As a consequence several attributes were pinpointed by the respondents, which have been included in the questionnaire.
- Computing statistics to calculate 'sample-size' for the main study.

h) Reliability of Instrument

The overall reliability of Cronbach's alpha was estimated at 0.9288, with 277 cases and 42 survey items. This value obviously exceeds the required minimum threshold for the overall Reliability-test, i.e. 0.7 (Koo, 2008).

IV. Analysis of the Empirical Data

a) *Descriptive Statistics*

Table 4.1 : Description of the Research Variables

Variables	Min	Max	Mean	Rank	Std. Deviation
HF	3.17	5.44	4.5559	4	.47526
OF	3.11	4.88	3.8071	5	.41125
TF	3.21	6.64	4.6851	3	.57352
ISD	3.50	5.45	4.7106	3	.46861
USE	2.13	6.21	5.6248	2	.78603
PRC	3.27	5.31	4.5234	1	.63711
PRB	2.23	5.11	4.3321	2	.56241
STF	2.47	5.39	4.5005	3	.77512
OPR	3.14	5.21	4.3101	2	.46327
SFF	2.22	6.11	5.5137	2	.67512

Table 4.2 : List of the Demographic Variables and Attributes

	Variables	Working Definitions (Attributes)	Code
1	Respondent-Type	Professors, Doctors, Consultants	RTP
2	Sector	Public and Private	PPR
3	Nature	Health/Education	HED
4	Gender	Male/Female	GDR
5	ICT-Background	IT People/Non-IT Workers	CNC
6	Age	Age of the Respondents	AGE
7	Experience	Using Computer Since	EXP
8	Designation	Designation of the Professors, Doctors and IT Consultants	DSG
9	City	Peshawar/Dera Ismail Khan	CTY

i. *Demographic Impacts*

The impacts of demographics on ISD and use practices are well documented by Wims & Lawler, 2007; Mehra & Mital, 2007. The developers of IT projects are constantly advised by the experts to address demographic differences regarding the development and use of IT projects for generating and sustaining positive user attitudes for effective uses of IT (Gay et al., 2006), which are based on the user-characteristics of gender, age, educational level, computer skills, experience with use of IT besides users styles, personal goals and attitudes, preferences, cultural background, experience, motivation (Moolman & Blignaut, 2008). The tables 4.3, 4.4 and 4.5 elaborate the statistics on demographic variables:

Table 4.3 : Type of Respondent, IT-people/Non-It Workers, Sector and Gender's Impacts

	Type of Respondent (df 2/351 = 3.0)		IT people/Non-IT Workers (df 352= 1.96)		Public/Private (df 352= 1.96)		Health//Education (df 352=1.96)		Gender (df 352= 1.96)	
Variables	F	p-Value	Cal. T-Val	p-Value	Cal. T-Val	p-Value	Cal. T-Value	p-Value	Cal. T-Val	p-Value
HF	5.417	.002	11.025	.000	-3.256	.002	11.024	.000	8.112	.000
OF	6.305	.001	10.946	.000	-3.829	.000	11.244	.000	4.235	.000
TF	26.032	.000	8.304	.000	-2.164	.018	9.404	.020	1.784	.050
ISD	.710	.331	12.556	.000	-4.873	.000	13.843	.000	5.822	.000
USE	25.374	.000	11.877	.000	-2.610	.006	14.565	.000	4.621	.000
PRC	10.230	.000	8.335	.000	-1.132	.207	10.351	.000	5.856	.000
PRB	12.111	.000	7.214	.000	-2.153	.017	12.240	.000	5.745	.000
STF	5.316	.001	10.021	.000	-4.762	.000	10.451	.000	5.711	.000
OPR	21.651	.000	10.835	.000	-3.651	.000	8.5313	.000	4.332	.000
SFF	22.263	.000	7.203	.000	-2.053	.017	8.338	.021	1.673	.040
	ANOVA		t-Test		t-Test		t-Test		t-Test	

Table 4.4 : Age, Experience and Qualification's Impacts

	Age (df 352= 1.96)		Exp with Computer (df 352= 1.96)		ICT-Q (df 352= 1.96)	
Variables	Cal. T-Val	p-Value	Cal. T-Val	p-Value	Cal. T-Val	p-Value
HF	-.204	.838	5.146	.000	7.271	.000
OF	-.129	.897	6.779	.000	9.513	.000
TF	1.219	.224	6.333	.000	5.691	.000
ISD	.127	.899	4.308	.000	12.742	.000
USE	-2.752	.006	5.363	.000	9.132	.000
PRC	.002	.998	6.012	.000	8.533	.000
PRB	1.331	.231	6.232	.000	4.580	.000
STF	-.201	.827	5.235	.000	6.161	.000
OPR	.133	.888	5.662	.000	8.402	.000
SFF	-211	.828	4.035	.000	6.160	.000
	t-Test		t-Test		t-Test	

Table 4.5 : City, Use of IT Since, Designation (Professor, Doctors and IT Consultants) Impacts

	City (df 352=1096)		Use of IT Since (df 352= 1.96)		Designation (Professors) (df 352= 3.0)		Designation (Doctors) (df 1/134=3.0)		Designation (IT Consultants) (df 352= 310)	
Variables	Cal T-Value	p-Value	Cal. T-Val	p-Value	F	p-Value	F	p-Value	F	p-Value
HF	-4.722	.000	-1.887	.460	-3.665	.002	.743	.920	.812	.710
OF	-3.446	.000	-2.055	.041	-.734	.842	2.488	.080	3.124	.011
TF	-.584	.377	-1.271	.157	1.264	.770	3.404	.000	1.424	.051
ISD	-1.610	.085	-3.041	.003	1.873	.233	.0239	.721	1.771	.021
USE	-3.641	.399	-1.666	.386	1.473	.366	.329	.730	1.521	.003
PRC	-4.030	.010	-1.244	.202	1.321	.273	1.351	.022	1.745	.061
PRB	-4.611	.000	-1.776	.461	-3.554	.001	.732	.911	.811	.711
STF	-3.335	.000	-2,043	.041	..635	.001	2.377	.070	3.013	.010
OPR	-1.434	.000	-2.154	.040	-3.556	.002	.732	.900	.701	.611
SFF	-.475	.000	-2.144	.156	1.153	.711	3.303	.000	1.346	.050
	t-test		t-Test		ANOVA		ANOVA		ANOVA	

b) Hypothesis Testing

Hypothesis No.1: The Public organizations are under-using IT potentials in comparison to private sector.

Results of independent sample t-test are shown in the below tables. As may be seen, the difference in the means of 3.65 and 2.58 with the standard deviations of .51 and .47 for the public and private respectively on the IT-potentials as IT use is significant. Similarly, calculated t value 14.234 in table No. 4.6 is greater than the tabulated t value 1.960, thus H_0 is not substantiated, which validates that public sector is under using IT potentials in comparison to private sector.

Group Statistics

	Nature	N	Mean	Std. Deviation	Std. Error Mean
IT Potentials	Public	149	3.6577	.518702	.03383
	Private	128	2.5812	.47114	.03541

Table 4.6 : Represents Groups Statistics for Hypothesis No. 1

a. Grouping Variables: Public, Private
b. Testing Variable: IT-Potentials

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
IT Potentials	Equal variances assumed	13.068	.000	14.324	398	.000	.87660	.06430	.86029	1.21311
	Equal variances not assumed			15.963	232.541	.000	.87660	.05817	.87210	1.10220

Table 4.7 : Represent the Results of Independent Sample t-test for Hypothesis No. 1

Hypothesis No.2: Escalation (time-delays, cost-overruns, compromise on lesser objectives) of IT projects is more common in public organizations than in private enterprises.

Below tables show the results of independent sample t-test for 2nd hypothesis. The difference in the means of 2.41 and 1.55 can be seen with the standard deviations of .47 and .31 for the public and private respectively for the escalatory behavior (time-delays, cost-overruns, compromise or lesser objectives for IT use is significant. As calculated t value 16.573 in table No. 4.9 is greater than the tabulated t value 1.960, thus H_0 is rejected.

Group Statistics

	Nature	n	Mean	Std. Deviation	Std. Error Mean
Escalation	Public	149	2.4174	.47104	.02817
	Private	128	1.5511	.31115	.02439

Table 4.8 : Represents Groups Statistics for Hypothesis No. 2

a. Grouping Variables: Public, Private
b. Testing Variable: Escalation

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Escalation	Equal variances assumed	38.919	.000	16.573	398	.000	.75712	.04891	.67099	.86330
	Equal variances not assumed			16.591	350.826	.000	.77623	.03726	.69387	.84042

Table 4.9 : Represent the Results of Independent Sample t-test for Hypothesis No. 2

Hypothesis No.3: IT-people overestimate while non-IT workers underestimate the role of IT in the organizations.

Results of independent sample t-test are shown in the below tables. As may be seen, the difference in the means of 3.14 and 2.03 with the standard deviations of .56 and .47 for the IT people and Non It workers respectively on the Role of IT in organizations is significant. As calculated t value .891 in table No. 4.11 is less than the tabulated t value 1.960, so H_0 hypothesis of the study is substantiated.

It can be inferred from the results that there is gap between IT people and Non IT workers with reference to role of IT in an organization which necessitates the education and intimate relations, corporation and coordination among these two groups

to development more understanding of the organization and management, technical competency and skills in their respective fields and to effectively use IT as competitive weapon for the accomplishment of organizational goals and objectives through innovation, growth, cost effectives, alliance and mergers.

Group Statistics

	User Types	N	Mean	Std. Deviation	Std. Error Mean
Role of IT in Org.	IT-People	149	3.1442	.56522	.04276
	Non-IT Workers	128	2.0333	.47731	.05241

Table 4.10 : Represents Group Statistics for Hypothesis No.3

a. Grouping Variables: IT-People, Non-IT Workers
b. Testing Variable: Role of IT in Org

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Role of IT in Org.	Equal variances assumed	.009	.924	.891	398	.550	.03878	.06700	-.10302	.16258
	Equal variances not assumed			.880	335.825	.550	.03878	.06770	-.10338	.16294

Table 4.11 : Represent the Results of Independent Sample t-test for Hypothesis No.3

Hypothesis No.4: Public sector is less optimistic about the role of IT than private sector.

Results of independent sample t-test for the fourth hypothesis are shown in the below tables. As may be seen, the difference in the means of 1.64 and 1.37 with the standard deviations of .45 and .34 for the public and private respectively on the Role of IT in organizations is significant. Where calculated t value 15.097 in table No. 4.13 is greater than the tabulated t value 1.960, Thus H_0 is not substantiated. This implies that private sector is more optimistic about the role of IT in organizations for maximum efficiency and effective utilization of both the human and material resources of the organization.

Group Statistics

	Nature	N	Mean	Std. Deviation	Std. Error Mean
Role of IT in Org.	Public	149	1.6437	.45367	.01545
	Private	128	1.3772	.37160	.02302

Table 4.12 : Show Group Statistics for Hypothesis No.4

a. Grouping Variables: Public, Private
b. Testing Variable: Role of IT

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Role of IT in Org.	Equal variances assumed	2.501	.115	15.097	398	.1400	.22762	.03662	.14397	.31827
	Equal variances not assumed			15.699	238.754	.1430	.23762	.04169	.15548	.31975

Table 4.13 : Represent the Results of Independent Sample t-test for Hypothesis No.4

Hypothesis No.5: Professors, doctors and consultants view IT differently.

The results of the 5th hypothesis are given in the below table. Since there are more than two groups and IT is measured on an interval scale, ANOVA is appropriate to test this hypothesis. If we look into the table, we find *df* in the 3rd column refers to the degrees of freedom, and each source of variation has associated degrees of freedom. For the between-groups variance, $df = (K-1)$, where K is the total number of groups or levels. Because there were three groups, we have (3-1) = 2 *df*. The *df* for the within groups sum of squares equals $(N-K)$, where N is the total number of respondents and K is the total number of groups. As there were no missing responses, the associated *df* is (277-3) = 276.

$$F = \frac{\text{MS explained}}{\text{MS residual}}$$

The mean square for each of variation (column 5 of the results) is derived by dividing the sum of squares by its associated *df*. Finally, the *F* value itself equals the explained mean square divided by the residual mean square. In this case, *F* = .240 (.014/.053). The *F* value is significant at the .676. As calculated *F* value .240 in table No. 4.14 is less than the tabulated *F* value 3.00, so H_0 hypothesis of this study is not substantiated. That is, there is no significant difference in the means implies that professors and doctors view IT differently from that of IT consultants.

ANOVA

IT

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.026	2	.014	.240	.676
Within Groups	11.146	204	.053		
Total	11.172	206			

Table 4.14 : Represents ANOVA Results for Hypothesis No.5

a. Grouping Variables: Professor, Doctors & IT consultants
b. Testing Variable: IT

Hypothesis No.6: Experience of non-IT workforce is negatively correlated with perceptions about IT.

Table 4.15 : Correlation

	HF	OF	TF	ISD	USE	PRC	PRB	STF	OPR	SFF	Average r
HF	1	0.651	0.44	0.611	0.746	0.486	0.404	0.409	0.541	0.301	0.535286
OF	0.651	1	0.758	0.746	0.834	0.732	0.349	0.455	0.632	0.647	0.646429
TF	0.44	0.758	1	0.577	0.745	0.665	0.334	0.334	0.466	0.403	0.550429
ISD	0.611	0.746	0.577	1	0.708	0.506	0.281	0.372	0.607	0.431	0.543
USE	0.746	0.834	0.745	0.708	1	0.718	0.719	0.431	0.617	0.734	0.700143
PRC	0.486	0.732	0.665	0.506	0.718	1	0.275	0.203	0.264	0.566	0.512143
PRB	0.404	0.349	0.334	0.281	0.719	0.275	1	0.263	0.348	0.271	0.375
STF	0.409	0.455	0.334	0.372	0.431	0.203	0.263	1	0.232	0.322	0.352429
OPR	0.541	0.632	0.466	0.607	0.617	0.264	0.348	0.232	1	0.305	0.524175
SFF	0.301	0.647	0.403	0.331	0.734	0.556	0.271	0.322	0.305	1	0.536238

Correlation is significant at the 0.01 level (2-tailed). (n=277)

Table 4.15 points correlations between the research variables, average correlations can be seen from last column. In the order of magnitude, the biggest weight of correlation is between the 'Satisfaction' and rest of the variables (r=0.7) and smallest correlation-score on Problems (r=0.512) and Development (r=0.535) with all the variables. However, 8 out of 10 variables are significantly correlated with r from 0.5, to 0.7.

Hypothesis No.7: Higher the perceptions about IT, greater are the chances/perceptions of success in IT projects.

On 5 point scale the relationship between higher perceptions about IT for greater chances/perception of success in IT projects was significant as tested by simple regression analysis. The first table lists the independent variable which is centered into the regression model and R (.104a) is the correlation of the independent variable with the dependent variable.

In the *Model Summery* table, The R Square (.011), which is the explained variance, is actually the square of the multiple R $(.104a)^2$. The *ANOVA* table shows that the *F* value of 4.217 is significant at the .038a. In the *df* (degree of freedom) in the same table, the first number represents the independent variable (1); the second number (277) is the total number of complete responses for all the variables in the equation (*N*), minus the number of independent variables (*K*) minus 1. (*N-K-1*) [(277-1-1) = 275]. The *F* statistic produced (*F = 4.217*) is significant at the .038a level.

To be statistically significant calculated correlation must be at least .304 on 5 point scale, it is inferred that the influence of perception about IT is significant as beta score is .513, thus H_0 hypothesis is not substantiated.

The next table titled *Coefficients* helps us to see that the independent variable influences most the variance in success of IT projects (i.e., is the most important). If we look at the column Beta under *Standardized Coefficients,* we see that the highest number in the beta for perception about IT .511 is significant at the .038a level. The results illustrate that the independent variable is significant.

This implies that perception about IT significantly influence the chances/perceptions of success in IT projects, thus the H_0 hypothesis is rejected.

Summary of Model

Model	R	R. Square	R. Square (Adjusted)	Estimation Std Error
1	.104(a)	.011	.008	.23501

a. Constant Predictors, Perception about IT, Success/Perception of IT Projects.

Table 4.16 : Represents Model Summary for Hypothesis No.7

ANOVA

Model		The sum of Squares	df	Square of Mean	F	Sig.
1	Regression	.239	1	.239	4.217	.038(a)
	Residual	21.982	275	.055		
	Total	22.220	277			

a. Constant Predictors, Perception about IT
b. Dependent Variable: Success/Perception of IT Projects

Table 4.17 : Represents ANOVA for Hypothesis No.7

Coefficients

Model		Non Standardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.128	.045		69.497	.000
	Perception about IT	.057	.028	.513	2.078	.038

a. Dependent Variable: Success/Perception of IT Projects

Table 4.18 : Show the Coefficients for Hypothesis No.7

Hypothesis No.8: The organizational, human, contextual and technological factors collectively determine the variation in the success/failure of an IT-project.

The multiple regressions analysis was applied according to standardized coefficient on 5 point scale for the dependence of success/failure of an IT-project on organizational, human, contextual and technology. The first table lists the four independent variables that are centered into the regression model and R (.561a) is the correlation of the four independent variables with the dependent variable, after all the intercorrelations among the four independent variables are taken into account.

In the *Model Summery* table, The R Square (.315), which is the explained variance, is actually the square of the multiple R $(.561a)^2$. The *ANOVA* table shows that the *F* value of 61.553 is significant at the .000a. In the *df* (degree of freedom) in the same table, the first number represents the number of independent variables (4); the second number (277) is the total number of complete responses for all the variables in the equation (*N*), minus the number of independent variables (*K*) minus 1. (*N-K-1*) [(277-4-1) = 272]. The *F* statistic produced (*F* = 61.553) is significant at the .000a level.

To be statistically significant calculated correlation must be at least 0.304 on 5 point scale, it is inferred that the influence of organization, human, context and technology on success of IT projects was found highly significant thus, the H_0 hypothesis is not substantiated.

The next table titled *Coefficients* helps us to see which among the three independent variables influences most the variance in success of IT projects (i.e., is the most important). If we look at the column Beta under *Standardized Coefficients,* we see that the highest number in the beta for organization .365, human .704, context .372 and for technology it is .533, which is significant at the .000a level. The results suggest the priority list for the policy makers to adopt it during the policy formulation for IT projects development and use process. It can be inferred that human play more important role than other factors i.e. organization, context and technology however, technology effects are greater than organizational and contextual factors.

Summary of Model

Model	R	R. Square	R. Square (Adjusted)	Estimation of Std Error
1	.561(a)	.315	.310	.19607

a. Constant Predictors, Organization, human, context and Technology.

Table 4.19 : Represents Model Summary for Hypothesis No.8

ANOVA

Model		The Sum of Squares	df	Square of Mean	F	Sig.
1	Regression	6.996	3	2.332	61.553	.000(a)
	Residual	15.224	396	.038		
	Total	22.220	399			

a. Constant Predictors, Organization, human, context & Technology.
b. Dependent Variable: Success of IT Projects

Table 4.20 : Result of ANOVA for Hypothesis No.8

Coefficients

Model		Non standardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std Error	Beta		
1	(Constant)	1.405	.137		10.243	.000
	Organization	.269	.043	.365	9.653	.000
	Human	.253	.029	.704	8.764	.000
	Context	.412	.054	.372	9.651	.000
	Technology	.468	.048	.533	9.782	.000

a. Dependent Variable: Success of IT Projects

Table 4.21 : Portray Coefficients for Hypothesis No.8

V. Major Findings and Discussion

Several studies have focused on the human challenges i.e. difference of perceptions about IT among the developers and users due to several gaps of education, communication, culture, motivation and satisfaction (Argyris, 1971; Kaasboll, 1997; Dann et al., 1998; Glass, 1998) while, Land et al., 1992; Ennals, 1995; Segars and Grover, 1996) studied the issues emanating from the organization and technology i.e. the nature, policies and procedures, the IT maturity, power structures etc. Likewise environment or context is significant because it influence and change altogether requirements for the success/failure of an IT project (Flowers, 1997). Similarly, Herzberg's two factors theory suggests that job-satisfiers relate to the job-contents while job-dissatisfiers emerge from the job-context (Luthans, 1995:149).

With this context, theoretical framework developed after literature review was used to get readings from the real-world situation (ISD and Use practices in KPK Pakistan). Primary data collected through questionnaire provided sufficient material about the problem-situation in the background of ideal theoretical framework extracted from the documented knowledge. The analysis and logical reasoning of the primary and secondary data provides good base for findings, following are the major findings along with discussion of this study:

The empirical results of this study points that public sector organizations in KPK Pakistan are less optimistic about the role of IT than private sector as indicated by the t value 15.097, which means that in KPK Pakistan, private sector is more optimistic about the role of IT in organizations for maximum efficiency and effective utilization of both the human and material resources of the organization that is why they are heavily investing in computerization of their organizational operations. This study further finds that public sector organizations are under-using IT potentials in comparison to private sector; the results of t statistics 14.234 support the literature. As for as Escalation in IT projects is concerned which are widely studied by researchers like Drummond (1994, 1996), again results of the study identified that escalation is severe issue of the public sector organizations than in private enterprises of KPK Pakistan according to t statistics 16.573. This implies that the ratio of time-delays, cost-overruns, compromise on lesser objectives is very high in public sector IT projects of KPK, which may leads to failure or total termination of projects, eating budget and resources of the organizations. Experts believe in application of soft methodologies and user participation in ISD (giving parallel importance to socio-technical factors) along with effective training and education of all the stakeholders involved also documented by Walsham (2000) Hirschheim and Klein (1989) Wynekoop and Russo (1995) Mumford and Weir (1979), however, this study have points that in comparison to private sector, public sector is ignoring these international signals and play down the human, social and psychological aspect in ISD, use and maintenance. The application of hard and fast rules with bureaucratic mind set (cumbersome procedures from project proposal to development, implementation and use are very common in public sector organizations. This may also result into miscommunication between the developer and user; make management of resistance to change more difficult.

The calculated *F* value .240 of this study explains the differences among professors and doctors and IT consultants who view ISD differently due their background diversities. Moreover the experience of non-IT workforce is negatively correlated with perceptions about IT that significantly affect the ISD and use process. Perceived ease of use, usefulness and experience with IT also play pivotal role in perception of users about IT projects. This study has found that IT-people overestimate while non-IT workers underestimate the role of IT in the organizations. This is verified by the t value .891 which pin point that there are gaps between IT people and Non IT workers with reference to role of IT in an organization which necessitates the education, close and intimate relations, corporation and coordination among these two groups to development more understanding of the organization and management, technical competency and skills in their respective fields and to effectively use IT as competitive weapon for the accomplishment of organizational goals and objectives through innovation, growth, cost effectiveness, alliance and mergers as higher the perceptions about IT, greater will be the chances/perceptions of success in IT projects. This is further supported by the Beta .511, which verified the arguments of Elton and Justin (1998) that higher perception about IT leads to greater success of IT projects development, use and implementation.

The nature (public/private), size, structure, objectives, and culture of the organization determine the organizational IT maturity i.e. the experience with ISD and use. The mechanism for developer-user interaction political/power struggles may help decrease the control the political maneuvering and powers struggle in IT projects development which according Sauer (1993, 1999), Markus & Bjorn-Andersen, 1987, Avgerou and Cornford (1998) and Glass (1998) is one of the major cause of IS projects failures.

The highest number in the beta of this study for organization .365 human .704, context .372 and for technology it is .533 supports the literature that organizational, human, contextual and technological factors collectively determine the variation in the success/failure of an IT-project and suggest the priority

list for the policy makers to devise strategies and policy when they are deciding about the IT projects development and use process. The beta .704 further highlight the value of that human element which play more important role than other factors i.e. organization, context and technology however, technology effects are greater than organizational and contextual factors. The common misperceptions about IT and perceptual gaps between developers and users as researcher's postulates with reference to ISD and use could be minimized through organizational motivation techniques for IT.

VI. Conclusions

The national IT policy is a very important document that set guidelines for the computerization in any country; Pakistan introduced its 1st IT policy in 1990 while Electronic Transaction Ordinance and Electronic Crimes Act were promulgated in 2002 and 2003 respectively, however according to Kundi (2009) there are several deficiencies and it is not comprehensive. Following are some suggestion for policy makers in the background of ISD and use practices in KPK Pakistan:

Promotion of IT-culture in all corners of the country among all segments of the society besides IT-education may be made compatible to the market needs, this demands revision of the old curricula and project management and evaluation techniques in IT education.

The feudal mind set of administrative machinery is also is the cause of failure of ISD and Use in KPK, so change in mind set of the administrative machinery and decision makers and effective training along with continuous updating of the information systems (eGovernment in particular) is required for effective ISD and use in KPK Pakistan. Moreover, human element play key role in success or failure of IT projects in comparison to technical factors, so developers are required not to ignore the human element rather give equal importance to socio-technical factors. Last but not the least is that administrative, socio-technical, political and cultural support is the backbone for successful development and implementation of IT project which must be ensured.

During the study it was observed that most of the IT projects were not completed within stipulated times which overburden the finances, inorder to remain economical and effective the project must completed within time and budget. The main reason of timely non completion is the political maneuvering and kickback involved in the projects besides imposing attitude and IT-organizational maturity which widens the gap between developers and users. In this connection Org-ware, people-ware, hardware and software training may be continuously provided to both developers and users, so that the common misperceptions about IT and perceptual gaps between developers and users may be minimized/ or bridged during the ISD and use for successful development and implementation.

Succinctly, one can understand that technology can be imported but not the demographic of the organization thus, non-technical issues are 'local in nature, structure and intensity,' which definitely need local studies of ISD and use practices so as to dig-out 'customized ISD and use process.

References Références Referencias

1. Argyris, C. (1971) Management information systems: The challenge to rationality and emotionality. Management Science. 17(6), 275-292.
2. Avgerou, C. and T. Cornford (1998) Developing information systems: concepts, issues and practice. 2nd ed. Macmillan Press Ltd.
3. Avison, DE. and AT. Wood-Harper (1990) MultiView: An exploration in information system development. Oxford. U. K. Blackwell Scientific Publications.
4. Avison, DE. and G. Fitzgerald (1995) Information system development: Methodologies, techniques, and tools. 2nd ed. McGraw-Hill Book Co.
5. Avison, DE. and H. Shah (1997) The information systems development life cycle: A first course in information systems. The McGraw-Hill Companies.
6. Babbie, E. (1993) the practice of social research. 7th ed. Wordsworth Publishing Co.
7. Baskerville, R. and S Smithson (1995) IT and new organizational forms: Choosing Chaos over panacea. European Journal of Information Systems. Vol.4, 66-73.
8. Beynon-Davies, P. (1995) Information systems 'failure': The case of London Ambulance Service's Computer Aided Dispatch project. European Journal of Information Systems. Vol. 4, 171-184.
9. Beynon-Davies, P. and M. Lloyd-Williams (1999) when health information systems fail. Health Informatics Journal. 19(4), 66-79.
10. Brooke, C. (1995) Analyst and Programmer Stereotypes: a Self-fulfilling prophecy? Journal of Information Technology. Vol. 10, 15-25.
11. Burn, JM. (1996) IS strategies and management of organizational change: A strategic alignment model. International Journal of Information Management. Vol. 8, 205-216.
12. Caudle, SL et al., (1991) Key information systems management issues for the public sector. MIS Quarterly/June. 171-188.
13. Collins, T. and D. Bicknell (1997) CRASH: Ten easy ways to avoid a computer disaster. Simon & Schuster A Viacom Company. UK.
14. Dann, J., A. Broome, and W. Joyce (1998) Human Barriers to Project Success the Computer Bulletin. May. 10-17.

15. DeMarco, T. (1979) Structured analysis and system specification. London. Englewood Cliffs, N.J.
16. Drummond, H. (1996) the politics of risk: trials and tribulations of the Taurus project. Journal of Information Technology. Vol. 11, 347-357.
17. Elton, J. and J. Roe (1998) Bringing discipline to project management. Harvard Business Review. March-April. 153-159.
18. Elton, J. and Justin, R. (1998) Bringing Discipline to Project management. Harvard Business Review, [Online] accessed from http://hbr.org/1998/03/bringing-discipline-to-project-management/ar/1
19. Ennals, R. (1995) Executive guide to preventing information technology disasters. Springer-Verlag London Ltd.
20. Ennals, R. (1995) Executive guide to preventing information technology disasters. Springer-Verlag London Ltd.
21. ETO (2002) Electronic Transaction Ordinance. Amendment of Articles 85, P.O No. 10 of 1984. Law Affairs Division, Ministry of Law and Legal Affairs, Government of Pakistan, Islamabad. 1-25.
22. Ewusi-Mensah, K. and ZH. Przasnyski (1991) On information systems project abandonment: An exploratory study of organizational practices. MIS Quarterly. March. 67-83.
23. Ewusi-Mensah, K. and ZH. Przasnyski (1991) on information systems project abandonment: An exploratory study of organizational practices. MIS Quarterly. March. 67-83.
24. Ewusi-Mensah, K. and ZH. Przasnyski (1994) Factors contributing to the abandonment of information System development Projects. Journal of Information Technology. Vol. 9, 185-201.
25. Ewusi-Mensah, K. and ZH. Przasnyski (1994) Factors contributing to the abandonment of information System development Projects. Journal of Information Technology. Vol. 9, 185-201.
26. Ewusi-Mensah, K. and ZH. Przasnyski (1995) Learning from abandoned Information Systems Development Projects. Journal of Information Technology. Vol. 5, 3-14.
27. Ewusi-Mensah, K. and ZH. Przasnyski (1995) Learning from abandoned Information Systems Development Projects. Journal of Information Technology. Vol. 5, 3-14.
28. Fitzgerald, B. (1996) Formalized system development methodologies: A critical perspective. Information Systems Journal. Vol. 6, 3-23.
29. Fitzgerald, B. (1996) Formalized system development methodologies: A critical perspective. Information Systems Journal. Vol. 6, 3-23.
30. Flowers, S. (1997) toward predicting information systems failure. DE. Avison (ed.) Key issues in information systems. Proceedings of the 2nd UKAIS Conference on University of Southampton. 2-4 April.
31. Galliers, RD. (1992) Information systems research, issues, methods and practical guidelines. Alfred Walker Ltd. Information System Series.
32. Glass, RL. (1998) Software Runaways: Lessons learned from massive software project failures. Prentice Hall and Yourdon Press.
33. Hirschheim, R., and HK. Klein (1989) four paradigms of information systems development. Communications of the ACM (32:10), 1989, pp. 1199-1215.
34. Kaasboll, JJ. (1997) How evolution of information systems may fail: many improvements adding up to negative effects: European Journal of Information Systems. Vol. 6, 172-180.
35. Koo, A. C. (2008). Factors affecting teachers' perceived readiness for online collaborative learning: A case study in Malaysia. Journal of Educational Technology & Society, 11 (1), 266-278. Retrieved November 10, 2008, from http://www.ask4research.info/
36. Land, FF., PN Le Quesne and I. Wijegunaratne, (1992) Technology transfer: Organizational factors affecting implementation, some preliminary findings. Cotterman, WW. and JA Senn (eds.) Challenges and strategies for research in system development. 65-81.
37. Markus, M. L. and Bjorn, A. (1987) Power over users: Its exercise by system professionals. Communications of the ACM. 27(12), 1218-1226.
38. Markus, ML. (1981) Implementation politics: Top management support and user involvement. Systems, Objectives, Solutions. Vol. 1. 203-215.
39. Markus, ML. (1983) Power, politics, and MIS implementation. Communications of the ACM. 26(6), 430-444.
40. McGrath, GM. (1997) an Implementation of Data-Centered Information Systems Strategy: A Power/Political Perspective. Failure and Lessons Learned in Information Technology Management. Vol. 1, 3-17.

Global Journal of Computer Science and Technology
Software & Data Engineering
Volume 12 Issue 15 Version 1.0 Year 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Encapsulation of Soft Computing Approaches within Itemset Mining – A Survey

By Jyothi Pillai & O.P.Vyas

Bhilai Institute of Technology, Durg, Chhattisgarh, India

Abstract - Data Mining discovers patterns and trends by extracting knowledge from large databases. Soft Computing techniques such as fuzzy logic, neural networks, genetic algorithms, rough sets, etc. aims to reveal the tolerance for imprecision and uncertainty for achieving tractability, robustness and low-cost solutions. Fuzzy Logic and Rough sets are suitable for handling different types of uncertainty. Neural networks provide good learning and generalization. Genetic algorithms provide efficient search algorithms for selecting a model, from mixed media data. Data mining refers to information extraction while soft computing is used for information processing. For effective knowledge discovery from large databases, both Soft Computing and Data Mining can be merged. Association rule mining (ARM) and Itemset mining focus on finding most frequent item sets and corresponding association rules, extracting rare itemsets including temporal and fuzzy concepts in discovered patterns. This survey paper explores the usage of soft computing approaches in itemset utility mining.

Keywords : *data mining, soft computing, itemset mining, fuzzy logic, neural networks, genetic algorithm.*

GJCST-C Classification : *H.2.8*

Strictly as per the compliance and regulations of:

Encapsulation of Soft Computing Approaches within Itemset Mining – A Survey

Jyothi Pillai [α] & O.P.Vyas [σ]

Abstract - Data Mining discovers patterns and trends by extracting knowledge from large databases. Soft Computing techniques such as fuzzy logic, neural networks, genetic algorithms, rough sets, etc. aims to reveal the tolerance for imprecision and uncertainty for achieving tractability, robustness and low-cost solutions. Fuzzy Logic and Rough sets are suitable for handling different types of uncertainty. Neural networks provide good learning and generalization. Genetic algorithms provide efficient search algorithms for selecting a model, from mixed media data. Data mining refers to information extraction while soft computing is used for information processing. For effective knowledge discovery from large databases, both Soft Computing and Data Mining can be merged.

Association rule mining (ARM) and Itemset mining focus on finding most frequent item sets and corresponding association rules, extracting rare itemsets including temporal and fuzzy concepts in discovered patterns.

This survey paper explores the usage of soft computing approaches in itemset utility mining.

Keywords : data mining, soft computing, itemset mining, fuzzy logic, neural networks, genetic algorithm.

I. Introduction

Association rule mining (ARM) is one of the most important areas of data mining research which is used for the discovery of frequent itemsets and their corresponding association rules[RT 2011]. An emerging topic in the field of data mining is Utility Mining which is an extension of Frequent Itemset mining. The main objective of Utility Mining is to identify the itemsets with highest utilities, by considering profit, quantity, cost or other user preferences. In many real-life applications, high-utility itemsets consist of rare items also[JV2010]. Soft computing aims to uncover the tolerance for vagueness, partial truth and approximation to achieve tractability, robustness and solutions with low cost. Soft computing methodologies consisting of fuzzy sets, neural networks, genetic algorithms, and rough sets are combined with data mining for knowledge discovery in large databases [SSP 2002]. The resultant technique is a more intelligent system which provides a human-interpretable, low cost solution.

This paper presents a brief overview of exploration of soft computing approaches in itemset utility mining. Section 2 and Section 3 discuss theoretical definitions related to Data Mining, Itemset Utility Mining and Temporal Mining. Section 4 discusses the state of art of soft computing tools. Section 5 presents usage of different soft computing methods in data mining, itemset mining and temporal mining. Section 6 presents conclusion and future work.

II. Data Mining

Data mining is the technique of automatic finding of hidden patterns and information elicitation from huge volume of raw data stored in data bases, data warehouses and other data repositories for making better business decisions, finding sales trends, in developing smarter marketing campaigns, and to predict customer loyalty.

Two categories of Data mining tasks are; Descriptive Mining and Predictive Mining. The Descriptive Mining techniques include Clustering, Association Rule Discovery, and Sequential Pattern Discovery, which is used to find human-interpretable patterns that describe the data in the form of clusters, itemsets, association rules and sequential patterns. The Predictive Mining techniques such as Classification, Regression, Deviation Detection, are used to classify objects or to predict future values of other variables.

One of the most important research areas in the field of Data mining is ARM. Association rules are used to identify relationships among a set of items in a transactional dataset. Apriori algorithm, given by Agrawal, Imielinski and Swami in 1993, is the first association rule mining algorithm, which influenced not only the association rule mining community, but also has impact on other data mining fields. Apriori and all its variants like Partition, Pincer-Search, Incremental, Border algorithm etc. take too much computer time to compute all the frequent item sets and usually consider only the frequency of items in itemsets.

III. Itemset Mining

a) Frequent Itemset Mining

Frequent itemsets are itemsets that occur frequently in a transaction data set. The goal of Frequent Itemset Mining is to identify all the frequent itemsets in a transaction dataset. A frequent itemset is the itemset having frequency support greater a minimum user specified threshold [JV2011].

Author α : Associate Professor, Bhilai Institute of Technology, Durg, Chhattisgarh, India. E-mail : jyothi_rpillai@rediffmail.com

Author σ : Professor, Indian Institute of Information Technology Allahabad, U.P., India. E-mail : dropvyas@gmail.com

Association Rule Mining (ARM)- The problem of mining association rules was first introduced in [RTA1993]. ARM is a popular technique for finding co-occurrences, correlations, frequent-patterns, associations among items in a set of transactions or a database. Rules with confidence and support above user-defined thresholds (minconf and minsup) were found. ARM process can be divided into two steps. The first step involves finding all frequent itemsets in databases. Next, association rules are generated from these frequent itemsets.

b) Rare Itemset Mining

The basic Bottleneck of ARM is Rare Item Problem. In many applications, some items appear very frequently in the data, while others rarely appear. In many practical situations such as security, business strategies, pattern extraction from web page access logs, biology, medicine and super market shelf management, the rare combinations of items in the itemset with high utilities provide very useful insights to the user [JV2010].

c) Utility Mining

Identification of the itemsets with high utilities is called as Utility Mining. The frequency of itemset is not sufficient to reflect the actual utility of an itemset [JV2011]. For example, the sales executives are not interested in frequent itemsets which do not yield significant profit. Mining of high utility itemsets is one of the most challenging recent data mining tasks. The utility value of an item depends on its evaluation e.g. if cola has support 30 and profit of 3%, cake may have support 10 but with a profit of 30%. This indicates that the utility of cake is higher than cola.

Utility mining model was proposed in [YHG2006] to define the utility of itemset. Utility is measured by analyzing how useful or profitable an itemset X is to user. The utility of an itemset X, u(X) is the sum of the utilities of itemset X in all the transactions containing X. An itemset X is called a high utility itemset if and only if u(X) >= min_utility, where min_utility is a user-defined minimum utility threshold [YHG2006]. For example, a computer system may be more profitable than a telephone in terms of profit. The main objective of high-utility itemset mining is to find all those itemsets having utility greater or equal to user-defined minimum utility threshold.

IV. Soft Computing

a) Importance of Soft Computing

Soft computing is tolerant of vagueness, imprecision, uncertainty, incomplete truth and approximation. The main components of Soft Computing are: fuzzy logic (FL), neural networks (NN), probabilistic reasoning (PR), genetic algorithms (GA), and chaos theory (ChT), which are summarized:-

i. *Fuzzy Logic* [RV 2011] Lotfi Zadeh conceived the concept of FL. FL is used to deal with uncertain or vague data, considered as fuzzy sets. In FL procedure, attribute values are transformed to fuzzy values and corresponding fuzzy membership or truth values are calculated.
ii. *Neural Networks* [RV 2011] NN is a network of artificial neurons which are simple processing elements which process information with a connectionist approach to computation [RAA2001]. An important property of these networks is their inductive nature, which uses "learning by example" in problems solving.
iii. *Genetic Algorithms* [RV 2011] GA is a flexible, heuristic and inductive search technique based on the theory of natural selection. GA learning consists of following steps: An initial input is created which consists of randomly generated rules. Each rule is represented using a string of bits. The fitness of a rule is evaluated by its classification accuracy on the training samples set. This process of generating new populations based on previous populations of rules is repeated till each rule of a population satisfies a pre defined fitness threshold [RAA2001].
iv. *Rough sets* [RV 2011] RS theory proposed by Pawlak is generally used for classification evaluation of data bases and for discovering structural relationships within uncertain or noisy data.
v. *Probabilistic Reasoning* PR [RAA2001]. PR offers methods to assess the outcome of systems which are affected by probabilistic ambiguity. The probabilistic mechanism provides a precise framework for illustration of a probabilistic knowledge, modeling of random phenomena and to analyze them.
vi. *Chaos Theory* [RAA2001]. A chaotic system is a deterministic system that exhibits random behavior. ChT deals with the non-linear dynamical systems that exhibit extreme sensitivity to initial conditions.

b) Need of Soft Computing in Data Mining

By incorporation of Soft Computing, there is a significant increase in effectiveness of artificial intelligence systems. All techniques have their own uniqueness based upon which they can be properly used in data mining process.

i. *Fuzzy Logic in Data mining*

The role of fuzzy sets based on different data mining functions are categorized below [SSP 2002]-

a. *Classification*

FL systems are used in several areas for classification, such as in business, health care and finance.

b. *Clustering*

Fuzzy clustering algorithms have been developed to mine telecommunications, customer and

prospect databases for gaining residential and business customer market share [SSP 2002].

c. *Association Rules*

Because of the affinity of FL with human knowledge representation, FL is considered as a key component of data mining systems.

d. *Functional Dependencies*

FL is also used for analyzing deductions based on functional dependencies (FDs) among variables, in database relations. Fuzzy relational databases generalize their classical and imprecise counterparts by supporting fuzzy information storage and retrieval [SSP 2002].

e. *Data Summarization*

FL techniques are used for data summarization.

ii. *Neural Networks in Data mining*

NNs can be efficiently encapsulated with data mining methods to increase the efficiency of the output of different data mining techniques. ANNs act as feasible computational models for different problems such as pattern classification, speech recognition, curve fitting, approximation capability, image data compression, associative memory, and modeling and control of non-linear unknown systems and are successfully utilized in various areas, such as science, engineering, medical, business, banking, telecommunication[RAA2001].

iii. *Genetic Algorithm in Data mining*

GA processing objects operate directly to set, queue, matrices, charts, and other structure. GA adopts probability rules to lead search direction. Genetic programming concepts have been used for developing Knowledge discovery systems. For better attribute interaction, GAs can be used.

iv. *Rough Sets in Data mining*

The main aim of RS is stimulation of approximation of concepts. Mathematical tools are offered by RS to extract hidden patterns in data and therefore are used in data mining. In data mining, RS can be used as a framework where precise data is not necessary and in the areas where approximate data is of great help. In data processing RST can be used for computing lower and upper approximation [RV 2011].

v. *Probabilistic Reasoning in Data mining*

Statistics or Probabilistic Theory forms a basis for good management and also plays a very important role in the data mining methods [RAA2001].

vi. *Chaos Theory in Data mining*

The predictability can be done using chaotic analysis and also prediction strategies of system's behavior can be formulated. ChT deals efficiently with noisy nonlinear systems. Chaotic computing gives a tool to determine a new perspective of nonlinear data analysis [RAA2001].

V. Literature Survey

a) Application of Soft Computing in Data Mining

By combining the advantages of both Data mining and soft computing paradigms, the techniques can be used for discovering knowledge in databases. In this section, a literature survey of integration of various soft computing methodologies and data mining is presented.

i. *Fuzzy Logic*

In retrieval of information, the main complexity is identifying relevant information, i.e. the nearest or the most similar according to user's need or expectation. This problem motivated to use fuzzy sets in knowledge representation thus enabling the user to express his prospect in a language not far from natural. Another reason is the approximate matching between the user's requirements and existing values in the database, on the basis of similarities and degrees of satisfiability.

The thesis report of Jianxiong Luo [JL1999] explores integrating FL with two data mining methods (association rules and frequency episodes) for intrusion detection. In intrusion detection, many quantitative features are involved and also security is fuzzy.

Au and Chan [WK1999] use an adjusted difference between experimental and probable frequency counts of attributes for finding out fuzzy association rules in relational datasets. The algorithm discovers both positive and negative rules and is able to cope with fuzzy class boundaries and missing values.

The authors in [BDLMR2007] focus on the applications of fuzzy techniques for information retrieval and data mining in real-world situations such as medical, educational, chemical and multimedia have been illustrated.

In real-time systems, for example in e-banking, assessing and determining any phishing websites is a complex and dynamic problem because of ambiguities involved. Aburrous et al present a intelligent, flexible and efficient system approach to deal with 'fuzziness' in the e-banking phishing website using fuzzy data mining techniques [AHDT2010].

In [KMA2012], the authors present an overview of the applications of fuzzy decision tree in heterogeneous fields. It is used dynamically in various fields such as intrusion detection, querying processes, cognitive process analysis (Human Computer Interaction), biometrics authentication, stock-market, parallel processing support, information retrieval and also in data mining.

ii. *Neural Network*

The paper [HRH1996] presents a method to find out symbolic classification rules using NNs.

Hongjun Lu et al propose an approach which can extract concise symbolic rules accurately using NN. The NN is trained for achieving required accuracy rate. Then through network pruning algorithm, repeated connections of the network are removed. The hidden layers of the network are analyzed and classification rules are generated and high quality rules are generated from the data sets.

In [X2008] the usage of NNs in data mining is researched in detail. NN can be considered as a parallel processing network which is formed by simulating the intuitive thinking of human. In data mining frequently used fuzzy NNs are fuzzy Back Propagation network, fuzzy perception model, fuzzy inference network, fuzzy clustering Kohonen network and fuzzy ART model.

By combining data mining and NNs, information is harvested from datasets by data warehousing firms [YA2009]. NNs can be used in all data mining tasks; generating association rules, classifications, clustering, prediction and forecasting. In data mining NNs act as a promising field for detecting and generating relationships among variables of large data sets.

NNs are motivated by brain functions, particularly pattern recognition and associative memory. The design of the NN architecture for the credit card detection system was based on unsupervised method, which was applied to the transactions data to generate four clusters of low, high, risky and high-risk clusters[F2011]. NN can be employed in banks to detect fraudulent usage of card more efficiently.

Anuj et al discuss that it is more expensive to connect a new customer than to maintain an existing loyal customer [AP2011]. The authors propose a NN based approach for predicting customer churn in cellular wireless services subscription and conclude a promising solution for customer churn management. The experimental results show that NN based method can predict customer churn with more than 92% accuracy.

Kamruzzaman et al propose a novel four-phase data mining algorithm using ANNs, referred as ESRNN (Extraction of Symbolic Rules from ANNs), for extracting symbolic rules [KJ2011]. The algorithm uses back propagation learning. Network architecture is defined and refined in the first phase and second phases. By using heuristic clustering algorithm, the nodes in hidden layers are discretized in third phase. Then symbolic rules are extracted from frequent patterns using extraction algorithm.

Mohammad Iquebal Akhter et al discusses in detail the function of ANN in preventing fraud in telecommunication services [MM2012]. A Fraud Detection System using ANN gathers historical data which is preprocessed and is used for training the NN for building a model which incorporates frequent fraud patterns. Finally, the model is applied to new business to learn new fraud patterns as the types of fraud evolved.

Madhusmita Swain et al introduced NNs for simplifying classification problem, IRIS plant classification [MSSA2012]. The problem identifies IRIS plant species on basis of plant attribute measurements. The authors used back propagation learning algorithm to train Multilayer feed- forward networks for identification of IRIS plants based on measurements such as length and width of sepal and length and width of petal. The authors conclude that Multi Layer Feed Forward NN (MLFF) is faster in terms of learning and is more accurate.

iii. *Genetic Algorithm*

A family of computational models which are inspired by evolution are GAs [E2011]. GA implementation begins with a population of random chromosomes. To create next generation of chromosomes from current population, GA uses three main types of rules:

1. The individuals called parents are selected through Selection rules, which contribute to next generation population.
2. Two parents are combined using Crossover rules to form next generation children.
3. Random changes are applied to individual parents using Mutation rules for forming children.

Ramesh Kumar et al presented a novel algorithm for rule prioritizing, which are generated by apriori algorithm through GA [RI2011].

E.P. Ephzibah proposed a new way to improve the performance of a model by combining GAs and FL, for feature selection and classification [E2011]. The proposed automated pattern classification system identifies and selects a subset of pattern from a larger set of features using fuzzy rule-based classification system. By the application of FL, the system's performance improved for diagnosing diabetes in patients.

Roohollah Etemadi et al propose a GA approach based on k-means clustering algorithm which can select cluster centers in a better manner [R2012]. All data objects are firstly clustered through k-means algorithm. Secondly, for each data object a pattern is generated by considering the generated clusters. On comparing with other related algorithms, the authors state that the proposed algorithm is more efficient than k-means algorithm and other algorithms.

Basheer M. Al-Maqaleh et al have explored the usage of GA, for finding predictive, complete and comprehensible classification rules from large database [BH2012]. The classification results of the proposed algorithm are compared with the performance of two algorithms; C4.5 and DTGA (DT and GA). DTGA has two rule inducing phases. In first phase, C4.5, a base classifier is used to generate rules from training data set,

then in next phase GA refines them for providing more accurate and high-performance prediction rules. According to authors, the proposed algorithm achieves better and accurate predictive results as compared to other two competent learners.

iv. *Rough Set*

RST deals with classificatory study of information systems. Z. Pawlak proposed this mathematical approach which is a powerful tool for dealing with vague data. Using RS method without deteriorating the quality of approximation, minimal attribute sets, and minimal length decision rules corresponding to lower or upper approximation can be extracted [W2012].

Prasanta et al proposed an approach based on RST which mine concise rules from inconsistent data [PRBB2011]. Firstly, lower and upper approximation is computed for each concept. Then a learning algorithm is adopted for building classification rules for each concept which satisfies classification accuracy. Test results show that the approach produced effective and minimal rules and offers more accurate results applied on several real life datasets.

In many fields such as inductive reasoning, classification, pattern recognition, cluster analysis, automatic learning algorithms, RST plays a significant role and is used in different domains like Medicine, Banking, Marketing and Engineering. In [S2011], A.S. Salama described some topological properties of RS which will help get rich results and discover hidden relations between data and also help in producing accurate programs.

Abdul Nassar proposes that using RST concept, clusters can be generated without any additional information for example probability distribution or fuzzy membership function [A2011]. By considering Lower approximation important rules of the target set can be generated. A reduct rule set of high importance can be generated by considering generated rules as attributes and a new decision table can be constructed.

Wen-Yau proposed a clustering technique which uses GA and RST [W2012]. After clustering, Apriori algorithm is used to discover association rules between products of same cluster and then marketing people can suggest related products to the targeting group. RS is used to generate rules and these rules are applied to various GA parts.

b) Application of Soft Computing in Itemset Mining

A literature survey of exploration of different soft computing approaches in itemset mining is discussed in this section.

i. *Fuzzy Logic*

Wai-Ho introduced a novel technique, called FARM (Fuzzy Association Rule Miner) to mine fuzzy association rules [WK1999] which uses linguistic terms for representing revealed regularities and exceptions, based on fuzzy set theory. The rules generated are called fuzzy association rules. FARM also discovers interesting associations between different quantitative values. One more advantage of FARM is that it can reveal both positive and negative association rules. A positive association rule indicates presence of another attribute value along with a certain attribute value whereas a negative association rule indicates absence of another attribute value along with a certain attribute value. Wai-Ho et al discuss that experimental results show FARM to be capable of discovering meaningful and useful fuzzy association rules.

Yi-Chung Hu et al proposed a learning algorithm, which acts as a knowledge acquisition tool for classification problems to efficiently generate fuzzy association rules [YRG2002]. In first phase, from training samples, large fuzzy grids are generated by fuzzy partitioning of each attribute and in second phase, for classification problems, fuzzy association rules by large fuzzy grids are generated. Experimental results on iris data indicate that the proposed algorithm accurately derive fuzzy association rules for classification problems.

One of the most essential areas of the application of fuzzy set theory is Fuzzy rule-based systems [CMM2004]. The advantages of using fuzzy systems for knowledge discovery processes are; information dealing with uncertain data, considering multi-variable relationships; human understandable results, easy information modification by an expert, easy adaptability to the given problem and high automated process. Fuzzy systems improve the interpretation and understandability of consumer models. In [CMM2004], Casillas et al presented a new approach for consumer behaviour modelling which is based on fuzzy association rules (FARs), centered on consumer attitude towards Internet and confidence in Internet shopping.

Sulaiman et al propose a new Fuzzy Healthy Association Rule Mining Algorithm (FHARM) which introduces new quality measures for generating more interesting and quality rules effectively and efficiently [SMCF2006]. Using FHARM, edible attributes are extracted from transactional input data and transformed to Required Daily Allowance (RDA) numeric values. The RDA values from database are then converted to fuzzy values. Analysis of normalized fuzzy transactional database is performed for getting nutritional information.

O. Dehzangi et al proposed a new approach to generate a set of rules for each class using data mining principles by reducing the number of generated rules [DZTF2007]. Using selection criteria, a precise number of rules for each class are selected and then a compact rule-base is constructed. The presented method improved the classification rate and deal effectively with noisy training examples.

Ashish Mangalampalli et al put forward a naive fuzzy ARM algorithm which performs faster and efficiently on very large datasets [AV2009]. Fuzzy ARM algorithm has following steps: Firstly, the crisp dataset is converted into a fuzzy dataset. Then fuzzy ARM algorithms are used which consider the fuzzy membership of an itemset in a given transaction along with its presence or absence.

Rajendran et al proposed a Novel Fuzzy Association Rule Mining (NFARM) method which deals with the detection of brain tumor in the CT scan brain images [RM2010]. In FARM, FL is used to transform numerical to fuzzy attributes. Discovered NFARM rules are tested on new test image to detect the brain tumor. The authors state that NFARM gives better performance and helps physicians in diagnosing the cancerous cells by providing better diagnosis system containing diagnosis keywords.

Radha et al proposed a classification method for generating fuzzy rules from training data [RR2010]. Using fuzzy C-Means algorithm, Quantitative attributes are divided into several fuzzy sets and accordingly membership values are generated. Then a supervised association rule algorithm is employed for discovering interesting FARs. Generated Fuzzy rules are used to build classification system. C4.5, Naïve Bayes, and ID3 classifiers are used for classification and accordingly fuzzy classified association rules are discovered. The authors discuss that the number of generated rules is reduced due to the usage of fuzzy linguistic values.

Maybin Muyeba et al presented a novel approach to mine weighted FARs effectively and address the issue of invalidation of downward closure property (DCP) in weighted ARM, where each item is assigned a weight according to its significance with respect to some user defined criteria [MSC2010].
Prakash et al present a qualitative fuzzy ARM (FARM) approach for mining FARs for the quantitative attributes [PP2011]. The authors evaluated the performance of qualitative FARM by experimenting with real data sets. Results prove that the qualitative approach discover more accurate association rules in less time with increased execution speed.

A novel approach is presented by Vedula Venkateswara Rao et al in [VES2012] for effectively mining frequent Item sets and generating association rules (ARs) based on fuzzy Apriori and weighted fuzzy Apriori. In weighted association rule mining (WARM), each item is assigned a weight with respect to its importance to some user defined criteria. Both binary data and fuzzy data are used in the proposed approach and Frequent Item Sets are generated. The Fuzzy Apriori algorithm (Apriori-Total) proposed in [VES2012] is founded on a tree structure called the T-tree to store frequent item set information.

K. Suriya Prabha et al proposed an approach that integrates FL and tree-based algorithm. The approach constructs a compact sub-tree for finding fuzzy frequent item [SL2012]. The authors conclude that the presented approach is quite efficient than other algorithms when evaluated on the basis of execution time, memory usages and search space for generating fuzzy frequent itemsets.

Ferdinando et al present a novel method for detecting association rules from datasets based on fuzzy transforms [FS2012]. AprioriGen algorithm is used for extracting fuzzy association rules which are represented in the form of linguistic expressions. A pre-processing phase is performed for determining optimal fuzzy partition of quantitative attributes domains.

Roohollah Etemadi states that one of the most well-known clustering methods is K-means algorithm which forms the base for other clustering approaches [R2012]. K-means and k-methods are heuristic partitioning algorithms where as Fuzzy k-means and Fuzzy k-methods are equivalent fuzzy type algorithms. In these partitioning methods, firstly k number of partitions is generated from data where each partition will contain at least one data. If crisp partitioning is performed, then a particular data will be present in only one cluster but if fuzzy partitioning is assumed then a particular data may be present in different clusters.

ii. *Neural Networks*

[VI2012] NNs have the capability to interpret meaning from complicated or vague data and hence can be used for extracting patterns and detecting trends which are difficult for humans or other computer techniques to notice

P. Sermswatsri et al proposes a more efficient method of frequent pattern mining by using Associative Classification method and NN [PS2006]. The proposed NN Associative Classification (NAC) method can be used to build more accurate and efficient classifiers. The authors conclude that the experimental results of NAC on datasets show improved accuracy rates.

Divya Bhatnagar et al propose an efficient technique for frequent itemsets mining in large databases using Optical NN Model [DNS2011]. The proposed technique removes the need for generating candidate sets for ARM to find frequent itemsets. The time complexity and space complexity of this technique is very low as optical NN can perform several computations simultaneously.

In [ASP2011], an efficient algorithm named Multi Level Feed Forward Mining (MLFM) is proposed by Amit Bhagat et al, for mining of multiple-level association rules efficiently from large transaction databases. The authors have used supervised NN in parallel for discovering frequent itemsets at each concept levels in a single scan of database. At each concept level MLFM reads items and divide them to various concept levels of hierarchy and passes it to the NN for generating frequent itemsets. Data at all the levels is given as input

by scanning the database only once, and thus produces fast output

NN Associative Classification system proposed by Prachitee B. Shekhawat builds a classifier with the help of Back propagation NN [PS2011].

iii. *Genetic Algorithm*

Xiaowei Yan et al designed an evolutionary mining strategy based on a GA called ARMGA model [XCS2007]. The authors discuss that, ARMGA model is efficient for global searching when search space is very large. Generally for rules mining, GAs are classified into two categories, according to encoding of rules in the population of chromosomes. In one encoding method called the Michigan Approach, each rule is encoded into an individual. In another method referred as the Pittsburgh Approach, the set of rules are encoded into a chromosome. ARMGA model is based on the Michigan strategy, where each association rule is encoded in a single chromosome.

In [ACC2007], Ansaf Salleb-Aouissi et al proposed QUANTMINER, a mining quantitative association rules system, which is based on GA that dynamically discovers "good" intervals in association rules.

Peter P. et al present a Pareto-based multiobjective evolutionary ARM method based on GAs [PV2008]. Predictive accuracy, comprehensibility and interestingness are used as different levels of interestingness of ARM problem

Xiaowei Yan et al designed a GA-based policy for discovering association rules without specifying actual minimum support [XCS2009].

Anandhavalli M. et al deal with a challenging ARM problem of finding optimized association rules [ASAG2009]. The authors by using GA find all the possible optimized rules from given data set. By using Apriori, frequent itemsets are generated.

Soumadip Ghosh et al propose a model in which the GA is applied on large data sets to find frequent itemsets [SSDP2010].

Vijaya Prakash et al proposed a technique to find all the frequent itemsets present in large data sets using GA [VGS2011]. The authors state that the generation of Frequent Itemset can be improved by using GA and also time complexity is reduced.

Rupesh Dewang et al propose "A new method for generating all positive and negative Association Rules" (NRGA) [RJ2011]. In First phase of NRGA, frequent itemsets and positive rules are generated using Apriori Algorithm. Then NRGA is used for generating all negative rules and finally GA is applied to optimize the generated rules.

Peter P. et al present a general ARM model for extracting useful information from very large databases [PVS2011]. The proposed model finds generalized association rules between items in a large database of transactions at any level of the taxonomy (is-a hierarchy) on the items.

The research paper [SJSK2012] presented by Sanat Jain et al is concerned with finding all positive and negative association rules from databases efficiently and optimization of generated rules is done using GA.

J. Malar Vizhi et al propose a new GA for generating high quality Association Rules. The authors used Michigan approach for representing the strong interesting association rules as chromosomes. Each chromosome is used to represent a separate strong association rule.

iv. *Rough Sets*

T. Y. Lin proposed a technique which applies RST to very large relational databases [T1996]. The proposed method integrates RST and the technique of extracting clean data subsets from noisy data banks for effectively mining soft rules.

Jiye Li et al introduced a rough set based process for ARM for selecting the most appropriate rules by using a rule importance measure [JN2005]. The authors introduced a rough set based model for providing an automatic and efficient way for ranking important rules in decision making applications.

X-Y. SHAO et al presents a methodology which integrates data mining tasks (like fuzzy clustering and ARM) and RST for discovering customer group-based configuration rules from the products purchased.

Tinghuai Ma et al provide a reduction algorithm for attribute reduction and pruning, using RST [TM2006]. The proposed reduction algorithm finds all reductions and is suitable for any uncertain knowledge reasoning.

In [EC2008], a new classification technique called 'Reduced MEPAR-miner Algorithm' is introduced by Emel Kizilkaya Aydogan et al, based on RST and Multi-Expression Programming for ARM, MEPAR-miner algorithm. In the preprocessing stage, the rough sets are used to reduce the feature space dimensionality and then to extract the classification rules, MEPAR-miner algorithms are used.

Anjana Pandey et al proposed RSMAR which uses Rough Set approach for Mining of Multidimensional Association Rules [AK2009]. The RSMAR algorithm consists of two steps. Firstly, the tables are combined to a single table for generating the rules which expresses the association between two or more domains belonging to different database tables and then on selected dimension, the mapping code is applied. In second step frequent itemsets are generated through equivalence classes and also the mapping code is transformed into real dimensions.

Jigyasa Bisaria et al have analyzed the sequential pattern mining problem through computational aspect and time constraint [JNP2009]. The authors have used RST to partition the sequential patterns search space in the proposed novel algorithm

which allows pre-visualization of patterns and also allows time constraint adjustment.

Anjana Pandey et al proposed an algorithm RS Model for Discovering Hybrid Association Rules [RSHAR] algorithm, for mining hybrid association rules using rough set approach [AP2009]. In RSHAR algorithm, the participant tables are combined into a general table for generating rules to express the relationship between two or more domains belonging to different database tables and then on selected dimension, the mapping code is applied. Then frequent itemsets are generated through equivalence classes and also the mapping code is transformed into real dimensions.

In [DHM2002], Daniel Delic et al emphasis on the comparison of association rules procedure and rough sets procedure. The proposed association rules method focus on the analysis of data bases containing boolean-valued attributes only. The authors conclude that there is a considerable reduction in computing time in the rough set algorithm.

A. Anitha et al proposed to combine upper approximation based rough set clustering and Apriori selective ARM for e-learning recommendation [AK2011]. In making e-learning recommendations, similar learning patterns are considered instead of all clicks stream sequences. The proposed algorithm resulted in dense clusters with less computational complexity and reduced number of extracted rules, which are highly relevant and meaningful.

VI. Conclusion

There has been substantial commercial interest as well as active research in data mining area for developing new and improved approaches for extracting information, relationships, and patterns from large datasets. Soft computing may be viewed as a foundation component for the emerging field of conceptual intelligence [RAA2001]. Hence Soft computing techniques can be encapsulated in Data mining for knowledge discovery in large databases. This paper presents a brief overview of various soft computing approaches used in itemset mining.

In future we will incorporate soft computing methodologies and itemset mining for mining high utility itemsets.

References Références Referencias

1. [A2011]Abdul Nassar . A.A, **Application of Rough Sets in Data Mining,** A Project Report Submitted in partial fulfillment of the requirements for the award of the degree of Master of Technology in Computer Science and Engineering, 2011.
2. [AHDT2010] M. Aburrous, M.A. Hossain, K. Dahal, F. Thabtah, **Intelligent phishing detection system for e-banking using fuzzy data mining,** Expert Systems with Applications 37 (2010), Elsevier, pp7913–7921.
3. [AK2011] A.Anitha, N.Krishnan, **A Dynamic Web Mining Framework for E-Learning Recommendations using Rough Sets and Association Rule Mining,** International Journal of Computer Applications (0975 – 8887), Volume 12– No.11, January 2011, pp 36-41.
4. [AP2009] Anjana Pandey, K.R.Pardasani, **Rough Set Model for Discovering Hybrid Association Rules,** IJCSIS June 2009 Issue, Vol. 2.
5. [AP2011] Anuj Sharma, Prabin Kumar Panigrahi, **A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services,** International Journal of Computer Applications (0975 – 8887), Volume 27– No.11, August 2011, pp 26-31.
6. [ASAG2009] Anandhavalli M., Suraj Kumar Sudhanshu, Ayush Kumar, Ghose M.K., **Optimized association rule mining using genetic algorithm,** Bioinfo Publications, Advances in Information Mining, ISSN: 0975–3265, Volume 1, Issue 2, 2009, pp 01-04.
7. [ASP2011] Amit Bhagat, Sanjay Sharma, K.R. Pardasani, **Ontological Frequent Patterns Mining by potential use of Neural Network,** International Journal of Computer Applications (0975–8887), Volume 36- No. 10, December 2011, pp 44-53.
8. [AV2009] Ashish Mangalampalli, Vikram Pudi, **Fuzzy Association Rule Mining Algorithm for Fast and Efficient Performance on Very Large Datasets,** Fuzzy Systems, FUZZ-IEEE 2009. IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Jeju Island, Korea, 20-24 Aug., 2009, pp 1163 – 1168.
9. [BDLMR2007] B. Bouchon-Meunier, M. Detyniecki, M.-J. Lesot, C. Marsala, M. Rifqi, **Real world fuzzy logic applications in data mining and information retrieval**. Springer, pp 219-247, 2007.
10. [BH2012] Basheer M. Al-Maqaleh , Hamid Shahbazkia, **A Genetic Algorithm for Discovering Classification Rules in Data Mining,** International Journal of Computer Applications (0975 – 8887), Volume 41– No.18, March 2012, pp 40 -44.
11. [CMM2004] J. Casillas, F.J. Martínez-López, F.J. Martínez, **Fuzzy Association Rules For Estimating Consumer Behaviour Models And Their Application To Explaining Trust In Internet Shopping,** Fuzzy Economic Review, Volume: 9, Nov. 2004, pp 3-26.
12. [DHM2002] Daniel Delic, Hans-J. Lenz, Mattis Neiling, **Rough Sets and Association Rules - which is efficient?,** 14th Conference on Computational Statistics (CompStat2002) Berlin, HU, August 24-28, 2002, pp 527- 532.
13. [DNS2011] Divya Bhatnagar, Neeru Adlakha, A. S. Saxena, **Mining Frequent Itemsets without Candidate Generation using Optical Neural Network**, IJCA Special Issue on "Artificial

Intelligence Techniques - Novel Approaches & Practical Applications", AIT, 2011, pp 14-18.

14. [DZTF2007] O. Dehzangi, M. J. Zolghadri, S. Taheri, S.M. Fakhrahmad, **Efficient fuzzy rule generation: A new approach using data mining principles and rule weighting,** Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007), August 24-August 27, ISBN: 0-7695-2874-0, Vol. 2, pp 134 – 139.
15. [E2011] E.P.Ephzibah, **Cost Effective Approach On Feature Selection Using Genetic Algorithms And Fuzzy Logic For Diabetes Diagnosis**, International Journal on Soft Computing (IJSC), Vol.2, No.1, February 2011, pp 1-10.
16. [EC2008] Emel Kizilkaya Aydogan , Cevriye Gencer, **Mining classification rules with Reduced MEPAR-miner Algorithm,** Applied Mathematics and Computation 195 (2008) pp 786–798, www.elsevier.com/locate/amc
17. [F2011] Francisca Nonyelum Ogwueleka**, Data Mining Application in Credit Card Fraud Detection System** Journal Of Engineering Science And Technology, Vol. 6, No. 3 (2011) 311 - 322.
18. [FS2012] Ferdinando Di Martino, Salvatore Sessa, **Detection of Fuzzy Association Rules by Fuzzy Transforms,** Advances in Fuzzy systems, 2012.
19. [HRH1996] Hongjun Lu, Rudy Setiono, Huan Liu, **Effective Data Mining Using Neural Networks,** IEEE Transactions On Knowledge And Data Engineering, VOL. 8, NO. 6, DECEMBER 1996, pp 957-961.
20. [JL1999] Jianxiong Luo, **Integrating Fuzzy Logic With Data Mining Methods For Intrusion Detection,** A Thesis Submitted to the Faculty of Mississippi State University in Partial Fulfillment of the Requirements for the Degree of Master of Science in Computer Science in the Department of Computer Science Mississippi State, Mississippi, August-1999.
21. [JN2005] Jiye Li, Nick Cercone, **A Rough Set Based Model to Rank the Importance of Association Rules,** In Proceedings of 10th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, RSFDGrC 2005, LNAI Volume 3642, pp 109-118.
22. [JNP2009] Jigyasa Bisaria, Namita Shrivastava, K.R. Pardasani, **A Rough Sets Partitioning Model for Mining Sequential Patterns with Time Constraint,** International Journal of Computer Science and Information Security(IJCSIS),Vol. 2, No. 1, 2009.
23. [JV2010] Jyothi Pillai, O.P. Vyas, **Overview of Itemset Utility Mining and its Applications,** International Journal of Computer Applications (0975 – 8887), Volume 5– No.11, August 2010.
24. [KJ2011] S. M. Kamruzzaman, A. M. Jehad Sarkar, **A New Data Mining Scheme Using Artificial Neural Networks,** Sensors 2011, 11, doi: 10.3390/s110504622, ISSN 1424-8220, www.mdpi.com/journal/sensors, pp 4622-4647.
25. [KMA2012] Kavita Sachdeva, Madasu Hanmandlu, Amioy Kumar, **Real Life Applications of Fuzzy Decision Tree,** Intl Journal of Computer Applications (0975–8887), Volume 42– No.10, March 2012,pp 24-28
26. [MB2012] J.Malar Vizhi, Dr. T.Bhuvaneswari, **Data Quality Measurement With Threshold Using Genetic Algorithm,** International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com, Vol. 2, Issue4, July-August 2012, pp 1197-1203.
27. [MM2012] Mohammad Iquebal Akhter, Mohammad Gulam Ahamad, **Detecting Telecommunication Fraud using Neural Networks through Data Mining**, International Journal of Scientific & Engineering Research IJSER, Volume 3, Issue 3, March-2012 1 ISSN 2229-5518, http://www.ijser.org, pp 1-5.
28. [MSC2010] Maybin Muyeba, M. Sulaiman Khan, Frans Coenen, **Effective Mining of Weighted Fuzzy Association Rules, 2010,** IGI Global, **pp 47-64,** DOI: 10.4018/978-1-60566-754-6.ch004**.**
29. [MSSA2012] Madhusmita Swain, Sanjit Kumar Dash, Sweta Dash, Ayeskanta Mohapatra, **An Approach For Iris Plant Classification Using Neural Network,** International Journal on Soft Computing (IJSC) Vol.3, No.1, February 2012, pp 79-89.
30. [PP2011] S.Prakash, R.M.S.Parvathi, **Qualitative Approach for Quantitative Association Rule Mining using Fuzzy Rule Set,** Journal of Computational Information Systems, http://www.Jofcis.com, 1553-9105, Binary Information Press, June2011,1879-1885.
31. [PRBB2011] Prasanta Gogoi, Ranjan Das, B Borah, D K Bhattacharyya, Efficient Rule Set Generation using Rough Set Theory for Classification of High Dimensional Data, International Journal of Smart Sensors and Ad Hoc Networks (IJSSAN) ISSN No. 2248-9738 (Print) Volume-1, Issue-2, 2011, pp 13-20.
32. [PS2006] P. Sermswatsri & C. Srisa-an, **A neural-networks associative classification method for association rule mining,** WIT Transactions on Information and Communication Technologies, Vol.37, pp 93-102, WIT Press, Southampton (2006).
33. [PS2011] Prachitee B. Shekhawat, Sheetal S. Dhande, **A Classification Technique using Associative Classification,** International Journal of Computer Applications (0975 – 8887), Volume 20– No.5, April 2011, pp 20-28.
34. [PV2008] Peter P. Wakabi-Waiswa, Venansius Baryamureeba, **Extraction of Interesting Association Rules Using Genetic Algorithms,** International Journal of Computing and ICT Research, Vol. 2, No. 1, pp. 26 – 33, June 2008, http:www.ijcir.org.
35. [PVS2011] Peter P. Wakabi-Waiswa, Venansius Baryamureeba, K. Sarukesi, **Generalized Association Rule Mining Using Genetic Algorithms,**

Computer Science, Proceedings of Seventh International conference in natural computation, IEEE 2011, 1116-1120, pp 59-69.

36. [R2012] Roohollah Etemadi, **A Novel Evolutionary Algorithm For Data Clustering In N Dimensional Space,** Indian Journal of Computer Science and Engineering (IJCSE), ISSN : 0976-5166, Vol. 2, No. 6, Dec 2011-Jan 2012, pp 902-908.
37. [RI2011] M. Ramesh Kumar, K. Iyakutti, **Application of Genetic algorithms for the prioritization of Association Rules,** IJCA Special Issue on "Artificial Intelligence Techniques - Novel Approaches & Practical Applications", AIT, 2011, pp 35-38.
38. [RJ2011] Rupesh Dewang, Jitendra Agarwal, **A New Method for Generating All Positive and Negative Association Rules,** International Journal on Computer Science and Engineering (IJCSE), ISSN: 0975-3397 Vol. 3, No. 4, Apr 2011, pp 1649- 1657.
39. [RM2010] P. Rajendran, M. Madheswaran, **Novel Fuzzy Association Rule Image Mining Algorithm for Medical Decision Support System**, International Journal of Computer Applications (0975 - 8887), Volume 1 – No. 20, 2010, pp 87-94.
40. [RR2010] R.Radha, S.P.Rajagopalan, **Generating Membership Values And Fuzzy Association Rules From Numerical Data,** (IJCSE) International Journal on Computer Science and Engineering, Vol. 02, No. 08, 2010, pp 2705-2715.
41. [RV2011] V. V. R. Raman, Veena Tewari, **Data-mining and Soft-computing: Basis for Technology-based Management,** International Conference on Technology and Business Management March 28-30, 2011, 1221-1226.
42. [S2011] A. S. Salama, **Some Topological Properties of Rough Sets with Tools for Data Mining**, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 2, May 2011, ISSN (Online): 1694-0814, www.IJCSI.org, pp 588-595.
43. [SCRPA2010] M. Sulaiman Khan, Frans Coenen, D. Reid, R. Patel, L. Archer, A sliding windows based dual support framework for discovering emerging trends from temporal data, Knowledge Based Systems, Volume 23, Number 4, May 2010, 316-322.
44. [SJSK2012] Sanat Jain, Swati Kabra, **Mining & Optimization of Association Rules Using Effective Algorithm,** International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, Volume 2, Issue 4, April 2012, pp 281-285, www.ijetae.com.
45. [SL2012] K. Suriya Prabha, R. Lawrance, **Mining Fuzzy Frequent itemset using Compact Frequent Pattern(CFP) tree Algorithm,** International Conference on Computing and Control Engineering (ICCCE 2012), 12-13 April, ISBN 978-1-4675-2248-9.
46. [SMCF2006] M. Sulaiman Khan, Maybin Muyeba, Christos Tjortjis, Frans Coenen, **An effective Fuzzy Healthy Association Rule Mining Algorithm (FHARM),** In Lecture Notes Computer Science, vol. 4224, pp.1014-1022, ISSN: 0302-9743, 2006.
47. [SSDP2010] Soumadip Ghosh, Sushanta Biswas, Debasree Sarkar, Partha Pratim Sarkar, **Mining Frequent Itemsets Using Genetic Algorithm**, International Journal of Artificial Intelligence & Applications (IJAIA), Vol.1, No.4, October 2010, pp 133-143.
48. [SSP 2002] Sushmita Mitra, Sankar K. Pal, Pabitra Mitra, **Data Mining in Soft Computing Framework: A Survey**, IEEE Transactions On Neural Networks, VOL. 13, NO. 1, JANUARY 2002, pp 3-14.
49. [T1996] T. Y. Lin, **Rough Set Theory in Very Large Databases,** Proceedings of Symposium on Modeling, Analysis and Simulation, IMACS Multiconference (Computational Engineering in Systems Applications, Lille, France July 9-12, 1996, Volume 2 of 2, 1095-1100.
50. [TM2006] Tinghuai Ma, Meili Tang, **A New Reduction Implementation Based on Concept,** Journal of Information and Computing Science, **ISSN** 1746-7659, England, UK, Vol. 1, No. 5, 2006, pp. 290-294.
51. [VGS2011] R. Vijaya Prakash, Govardhan, S. S.V.N. Sharma, **Mining Frequent Itemsets from Large Data Sets using Genetic Algorithms,** IJCA Special Issue on "Artificial Intelligence Techniques - Novel Approaches & Practical Applications", AIT, 2011, pp 38-43.
52. [VI2012] Vinodhini Katiyaar, Ina Kapoor Sharma **Use of Data Mining & Neural Network in Commercial Application** International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 3 (3) , 2012,ISSN 0975-9646, pp 4041 – 4049.
53. [W2012] Wen-Yau Liang, **A Hybrid Evolutionary Computing For Market Segmentation,** Business and Information 2012, Sapporo, July 3-5), pp F171-F177.
54. [X2008] Xianjun Ni, **Research of Data Mining Based on Neural Networks**,World Academy of Science, Engineering and Technology 39 2008.
55. [XCS2007] Xiaowei Yan, Chengqi Zhang, Shichao Zhang, **ARMGA: Identifying Interesting Association Rules With Genetic Algorithms**, Applied Artificial Intelligence, 19:677–689, 2007, ISSN: 0883-9514 print/1087-6545 online, pp 677-689.
56. [XCS2009] Xiaowei Yan , Chengqi Zhang , Shichao Zhang , **Genetic algorithm-based strategy for identifying association rules without specifying actual minimum support,** Expert Systems with Applications 36 (2009) 3066–3076, www.sciencedirect.com.

57. [YRG2002] Yi-Chung Hu, Ruey-Shun Chen, Gwo-Hshiung Tzeng,**Mining fuzzy association rules for classification problems,** Computers & Industrial Engineering 43 (2002), Elsevier Science Ltd., pp 735–750.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Software & Data Engineering
Volume 12 Issue 15 Version 1.0 Year 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

E-learning Opportunities & Prospects in Higher Education Institutions of Khyber Pakhtunkhwa, Pakistan

By Ghulam Muhammad Kundi & Allah Nawaz

Gomal University Dera Ismail Khan, Pakistan

Abstract - Both opportunities and prospects are sometimes used interchangeably however, in this paper, opportunity refers to the 'availability of eLearning resources and service' while prospects denote 'futuristic expectations about the role of information and communication technologies (ICTs) in higher education institutions (HEIs). The empirical findings suggest that people score lower on opportunities but significantly high on the prospects showing that they are not quite happy with the facilities and services available (due to the development, implementation and use problems – or simply management problems of eLearning). But they can clearly foresee the significant role of ICTs or education technologies (ETs) in future in the context of developing countries like Pakistan. Furthermore, these differences are attributed to the demographic diversities of the respondents, meaning that the demographic variation changes the power and direction of the user-attitudes towards eLearning. This paper uses stepwise regression to gradually glean-out the most significant predictors of opportunities and prospects from a group (eight) of demographics.

Keywords : *ICTS, ETS, HEIS, VLE, elearning, eteaching, epedagogy, ecourses, opportunities, prospects, demographic-attributes.*

GJCST-C Classification : *J.1*

Strictly as per the compliance and regulations of:

E-learning Opportunities & Prospects in Higher Education Institutions of Khyber Pakhtunkhwa, Pakistan

Ghulam Muhammad Kundi [α] & Allah Nawaz [σ]

Abstract - Both opportunities and prospects are sometimes used interchangeably however, in this paper, opportunity refers to the 'availability of eLearning resources and service' while prospects denote 'futuristic expectations about the role of information and communication technologies (ICTs) in higher education institutions (HEIs). The empirical findings suggest that people score lower on opportunities but significantly high on the prospects showing that they are not quite happy with the facilities and services available (due to the development, implementation and use problems – or simply management problems of eLearning). But they can clearly foresee the significant role of ICTs or education technologies (ETs) in future in the context of developing countries like Pakistan. Furthermore, these differences are attributed to the demographic diversities of the respondents, meaning that the demographic variation changes the power and direction of the user-attitudes towards eLearning. This paper uses stepwise regression to gradually glean-out the most significant predictors of opportunities and prospects from a group (eight) of demographics.

Keywords : ICTS, ETS, HEIS, VLE, elearning, eteaching, epedagogy, ecourses, opportunities, prospects, demographic-attributes.

I. Introduciton

Opportunities are the user-perceived benefits in ICTs while Prospects refer to the perceived future of ETs or eLearning tools in higher education. The opportunities and particularly, prospects are very highly scored around the world. Teachers, students and administrators are very positive about the existing opportunities provided by the ICTs and the future of these technologies in higher education. Even when many problems are reported by the respondents with regard to the installation and use of eLearning systems, they score high on the opportunities and prospects showing that despite the problems, ICTs have the future. It also shows that users believe in the opportunities conceived in these technologies but there are problems in their management and use.

The current trend in eLearning ventures is collaborative development and operation. The researchers have documented volumes of research suggesting that if eLearning is build more according to the contextual demands, there are brighter chances of a successful effort (Chan & Lee, 2007). Traditionally, 'one-for-all' model has prevailed, which did not appear as a good option in many situations thereby opening research about the contextual determinants of eLearning projects. Researcher over research has confirmed that compatibility of new tools with user-demographics and environmental dimensions are the only criteria for future eProjects of eLearning in HEIs (Nawaz & Kundi, 2010a).

This gap is indicative of the problems and obstacles which are holding back the university constituents to fully integrate ICTs in their teaching, learning and administrative functions. These barriers come from the user-demographics and the factors concerning eLearning-environments in HEIs, such as, ETs, Development and Use practices, and User Training and Satisfaction etc, meaning that the gap is between the 'user and environmental-requirements' and 'whatever is available to the users in practice – the contextual mismatch' (Nawaz et al., 2007; Qureshi et al., 2009; Nawaz & Kundi, 2010b). This paper is an effort to study the stepwise regression to gradually glean-out the most significant predictors of opportunities and prospects from a group (eight) of demographics in HEIs of Khyber Pakhtoonkhwa, Pakistan.

II. Literature Review

ICTs are providing several opportunities to all the countries of the world thereby creating the brighter prospects of eLearning particularly for the developing states in handling their long-standing problems of mass education (Tinio, 2002; Oliver, 2002). ICTs are capable to increase the opportunities of active learning, inter-connectivity, enhanced feedback (Abrami et al., 2006) and a working environment of teamwork and collaboration (Chan & Lee, 2007). Views of the eLearning-users are founded on their 'digital-literacy' which builds their attitudes towards ICTs, ETs and eLearning in higher education (Kundi & Nawaz, 2010) as well their demographic attributes (Nawaz & Kundi, 2010a).

a) Opportunities of eLearning

A repeated claim of the technology-proponents is that ICTs conceive unprecedented opportunities,

Author α σ : Assistant Professo Department of Public Administration Gomal University Dera Ismail Khan, Pakistan.
E-mails : kundi@gu.edu.pk, profallahnawaz@gmail.com

particularly, for the 'developing-countries'. This optimism is founded on the premise that the miraculous capabilities of the digital-gadgets have transformed the society into a 'global-village' through a kind of connectivity, which is never quoted in the history of mankind (Nawaz et al., 2007). UNESCO (2007) reports that the use of ICTs in and for education is rapidly expanding in many countries and considered both as a necessity and an opportunity. Research also suggests that ICTs offer new learning opportunities for students (eLearning), develop teacher's professional capabilities (ePedagogy) and strengthen institutional capacity (eEducation) (Ezziane, 2007) and most universities today offer some form of eLearning (Kanuka, 2007).

Virtual learning environments (VLEs) have emerged with tools and techniques for the course-management and interactivity of teachers and learners through a long line of opportunities particularly, the web-based applications, which enable not to simply deliver knowledge rather empower learners to develop research skills and capitalize on web to "harvest knowledge (Gray et al., 2003)." Similarly, Internet offers opportunities which need to be explored, the technologies are designed well and used as intended (Wijekumar, 2005). Thus, eLearning offers a "great and exciting opportunities for both educators and learners (Manochehr, 2007)."

One of big expectations tied to e-learning speaks about its ability to introduce equal education to everyone. Authors of this assert that the possibility of e-courses to reach any corner of our planet will lead to the opportunity of delivering same high-quality education everywhere. The biggest optimists have a vision of top-ranking universities acting over the Internet using ready-made courses for huge amounts of students in Third-World countries. In accordance to well-known practices of e-learning, the students would study on their own pace by self-learning (Hvorecký, 2005). Because e-learning is supported by internet and web technologies, which are delivered via end-user computing that creates connectivity between people and information, and offers opportunities for social learning approaches (Luck & Norton, 2005). For example, a new feature of eLearning 'Blogs' provide the opportunity for feedback from anyone in the world creating limitless collaborative options. Succinctly, they are potentially powerful collaborative tools to build writing ability (Drexler et al., 2007).

New technologies reduce transaction costs for reproduction and distribution to a minimum. In principle, ICTs offer the opportunity to merge two formerly distinct processes, publishing and archiving, into one integrated activity. To put a document in an online repository is simultaneously a step to publish it. Without covering the full range of possibilities, we discuss three different types: self-archives online-journals and pre-print-servers (Pfeffer, 2004). As we entered into the third millennium, education via internet, intranet or network represents great and exciting opportunities for both educators and learners (Manochehr, 2007). While instructors cannot always accommodate each student's need, it is important that several learning opportunities are provided (Manochehr, Naser-Nick (2007)).

b) Prospects of eLearning

Universities are challenged to integrate ICTs into their strategies, their institutions and educational processes. Policy responses are better if devised at national and supranational levels, the major aims being the improvement of quality and flexibility, the widening access to the field of tuition, the possibility of reaching populations as yet un-reached by higher education. Such missions are those of the "Mega-Universities", those large distance education institutions which are already broadening the scope of higher education in several countries. When ICTs are adapted to local technological conditions, they become a major tool both for on-campus students, and for reaching the new target groups engaged in lifelong learning processes or on professional markets (Loing, 2005).

Researchers predict the prospects of 'multi-versities' focusing on the provision of a large diversity of programs, and 'flexi-versities' featuring market specialization and staff and student flexibility. This change in the universities represents a move "from being scholarly ivory towers to information corporations (UQA, 2001)." Thus, ICTs have prospects for universities in developing countries to improve their teaching and learning processes. It is argued that, universities in developing countries should adopt eLearning technologies to improve teaching and leaning processes. Pedagogical, technical and cost issues should be taken into account for each specific technology when integrating ICTs in teaching and learning practices (Sife et al., 2007).

ICT-based education is seen as "the dominant engine for productivity improvement and business opportunities" and "a key factor for generating future employment"(Hagan, 2003). For instance virtual or distance learning can help to overcome the problems associated with geographical isolation and is invaluable for students in remote areas.

Distance learning educational software also benefits from economies of scale increasing cost efficiencies. Recruiting teachers for the more remote regions is often difficult in Developing Countries; ICT serves to counteract physical distance as teachers can maintain contact with family and friends through telephone and e-mail (Wims & Lawler, 2007). However, to increase the prospects of eLearning to improve higher education requires reshaping of the mindset and practices in the teaching, learning and educational administration (Thompson, 2007; Qureshi et al., 2009; Kundi & Nawaz, 2010).

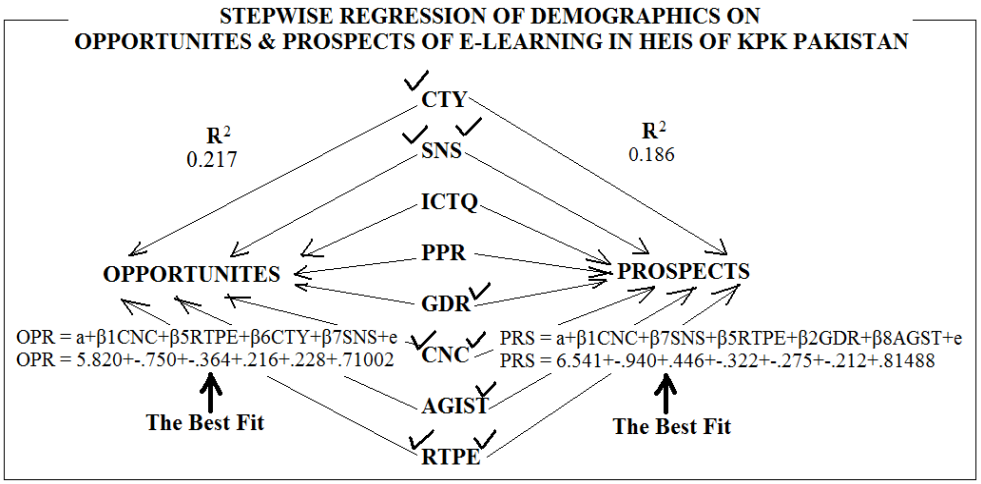
c) Demographic Implications

Research shows that despite the claimed advantages of eLearning, problems can arise if new systems are not compatible with the learner characteristics like nationality and gender (Graff et al., 2001). Although, with regard to an individual user, two key factors are users' motivation towards eLearning and their capabilities in using eLearning facilities (Lynch et al., 2005) however, the users' attitude towards ETs depends on their personal characteristics including age, gender, teacher-centric vs. student-focused teaching and learning, digital literacy, and learning styles (Cagiltay et al., 2006). Other researchers support this idea by noting that teachers' use of ICTs is influenced by the factors like: demographic-attributes (age, educational background etc); access to hardware; experience in using computers and perceptions about the usefulness and ease of using new digital gadgets (Mehra & Mital, 2007).

Thus, the demographic impacts on user perceptions, theories, and attitudes on the development and use of eLearning in HEIs are well documented (Valcke, 2004; Gay et al., 2006; Wims & Lawler, 2007). The developers of eLearning systems are repeatedly advised to address demographic differences through devising such strategies, which generate and sustain positive attitudes of users in eLearning environments (Gay et al., 2006). These differences emanate from the user-characteristics of gender, age, educational level, computer skills, previous experience with eLearning, learning styles, personal goals and attitudes, preferences, cultural background and motivation (Moolman & Blignaut, 2008; Nawaz & Kundi, 2010a).

Figure 1 portrays a graph of the theoretical model showing the structure and distribution of the hypothesis tested for this publication and empirical outputs computed through stepwise regression analysis. Both R^2 and the best-fit models have also been given in the figure.

Figure 1 : Schematic Diagram of the Theoretical Framework



III. Research Design

Survey approach has been used in this project by selecting a sample from the population of teachers, students and administrators in the higher education of the KPK. Population of this study includes all the HEIs in the province while sample included all the institutions in two cities of Peshawar and Dera Ismail Khan (DIK) (big & small cities respectively), selected due to the following features:

a. Peshawar (big city) and Dera Ismail Khan (DIK) (small city).
b. Both the cities host two of the oldest universities of the province (University of Peshawar – 1950 and Gomal University - 1974).
c. The cities have both the oldest as well as new universities (pre-2000 and the post-2000) working in public and private sectors.
d. These institutions are populated with students, teachers and administrators from almost all cities and areas of the province.

A structured questionnaire was developed from the existing literature by extracting both research and demographic variables. Besides demographics, the variables were about the perceptions of users about educational technologies, their available opportunities and expectations of the students, teachers and administrators about the future prospects of eLearning in HEIs (30 items on 7-point scale). The questions relating to the available opportunities and future prospects were 9 and 7 respectively. The Cronbash's alpha was estimated at 0.9288, with 354 cases and 38 survey items (with eight demographics). This value is acceptable as it exceeds the required minimum score of 0.7 for overall reliability (Koo, 2008).

We used SPSS 12.0 to create the database for applying statistical procedures to produce descriptive tables and test the hypotheses for inferential analysis. For testing of hypotheses, stepwise multiple regression procedures was used to gradually eliminate the weak predictors from the 'best-fit' for the prediction of opportunities and prospects. Two research-variables (Current Opportunities and Future Prospects of eLearning) were selected for computing the impacts of eight demographics on the respondents' attitudes. All the demographic-attributes were converted into 'Dummy-variables' with 0 and 1 as codes for all the variables.

IV. Findings Of The Study

a) *Demographic Groups*

Table 1 : Frequencies of the Demographic Groupings (n=354)

1	**City - CTY**	**Frequency**	**Percent**	**Valid Percent**
	Small City (D. I. Khan)	145	41.0	41.0
	Big City (Peshawar)	209	59.0	59.0
2	**Science/Non-Science - SNS**			
	Science Respondents	152	42.9	42.9
	Non-Science Respondents	202	57.1	57.1
3	**ICT Qualification - ICTQ**			
	Formal Computer Qualification	119	33.6	33.6
	Informal Computer Qualification	235	66.4	66.4
4	**Public/Private - PPR**			
	Public Universities	180	50.8	50.8
	Private Universities	174	49.2	49.2
5	**Gender - GDR**			
	Male Respondents	241	68.1	68.1
	Female Respondents	113	31.9	31.9
6	**Computer/Non-Computer - CNC**			
	Computer (as a Subject)	101	28.5	28.5
	Non-Computer (other Subjects)	253	71.5	71.5
7	**Age of the Institute - AGIST**			
	Pre2000 (established before 2000)	191	54.0	54.0
	Post2000 (established after 2000)	163	46.0	46.0
8	**Respondent-Type - RTPE**			
	Student Respondents	132	37.3	37.3
	Teachers & Administrators	222	62.7	62.7

b) *Regression of Demographics on Opportunities of eLearning*

i. *Models, Coefficients & Excluded Variables* (OPR)

Table 2 : Showing the Details of the FOUR Models

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	F	Sig.
1	.376(a)	.141	.139	.74047	57.803	.000(a)
2	.430(b)	.185	.180	.72242	39.768	.000(b)
3	.452(c)	.205	.198	.71461	29.999	.000(c)
4	.466(d)	.217	.208	.71002	24.177	.000(d)
Detail of the Models	a Predictors: (Constant), CNC b Predictors: (Constant), CNC, RTPE c Predictors: (Constant), CNC, RTPE, CTY d Predictors: (Constant), CNC, RTPE, CTY, SNS e. Dependent Variable: OPPORTUNITES					

Table 3 : Showing the Coefficients of Regression in FOUR Models

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.787	.074		78.544	.000
	CNC	-.663	.087	-.376	-7.603	.000
2	(Constant)	5.982	.085		70.555	.000
	CNC	-.632	.085	-.358	-7.411	.000
	RTPE	-.346	.080	-.210	-4.337	.000
3	(Constant)	5.826	.099		58.779	.000
	CNC	-.595	.085	-.337	-6.972	.000
	RTPE	-.357	.079	-.216	-4.519	.000
	CTY	.231	.078	.142	2.952	.003
4	(Constant)	5.820	.099		59.081	.000
	CNC	-.750	.107	-.425	-6.982	.000
	RTPE	-.364	.078	-.221	-4.644	.000
	CTY	.216	.078	.134	2.777	.006
	SNS	.228	.097	.142	2.354	.019

Dependent Variable: Opportunities of eLearning in HEIs of KPK, Pakistan

Table 4 : Showing the Excluded Variables in FOUR Models

Model		Beta	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
4	ICTQ	.028(d)	.327	.744	.018	.307
	PPR	-.077(d)	-1.527	.128	-.082	.881
	GDR	-.033(d)	-.660	.510	-.035	.926
	AGIST	-.015(d)	-.320	.749	-.017	.965

ii. *Analysis I*

Regression models in table 2 gives the detail of all four procedures applied to find the best fit equation to predict the opportunities of eLearning as expressed by the respondents with differing demographic features. As given in the table, first model explains 14% of the variation in opportunities however as the new models are developed the percentage goes up and ultimately, fourth model predicts 22% of the dependent variable. Similarly, table 4 gives a list of excluded variables with p-values greater than the required .05 to test the hypotheses.

The best fit equation is:

$OPR = a+\beta_{1CNC}+\beta_{5RTPE}+\beta_{6CTY}+\beta_{7SNS}+e$

$OPR = 5.820+-.750+-.364+.216+.228+.71002$

c) Regression of Demographics on Prospects of eLearning

i. *Models, Coefficients & Excluded Variables* (PRS)

Table 5 : Showing Coefficients of Regression in FIVE Models

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	F	Sig.
1	.329(a)	.109	.106	.84816	42.860	.000(a)
2	.369(b)	.136	.131	.83603	27.702	.000(b)
3	.394(c)	.155	.148	.82810	21.408	.000(c)
4	.416(d)	.173	.164	.82043	18.252	.000(d)
5	.432(e)	.186	.175	.81488	15.955	.000(e)
Detail of the Models	a Predictors in the Model: (Constant), CNC b Predictors in the Model: (Constant), CNC, SNS c Predictors in the Model: (Constant), CNC, SNS, RTPE d Predictors in the Model: (Constant), CNC, SNS, RTPE, GDR e Predictors in the Model: (Constant), CNC, SNS, RTPE, GDR, AGIST f Dependent Variable: PRC_PRS					

Table 6 : Showing Coefficients of Regression in FIVE Models

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	6.203	.084		73.499	.000
	CNC	-.654	.100	-.329	-6.547	.000
2	(Constant)	6.169	.084		73.611	.000
	CNC	-.911	.125	-.459	-7.304	.000
	SNS	.382	.114	.211	3.360	.001
3	(Constant)	6.311	.097		64.735	.000
	CNC	-.899	.124	-.453	-7.267	.000
	SNS	.397	.113	.219	3.516	.000
	RTPE	-.255	.091	-.137	-2.785	.006
4	(Constant)	6.421	.105		61.430	.000
	CNC	-.888	.123	-.448	-7.249	.000
	SNS	.414	.112	.229	3.695	.000
	RTPE	-.321	.094	-.173	-3.424	.001
	GDR	-.267	.097	-.139	-2.752	.006
5	(Constant)	6.541	.115		56.780	.000
	CNC	-.940	.124	-.474	-7.606	.000
	SNS	.446	.112	.246	3.981	.000
	RTPE	-.322	.093	-.174	-3.462	.001
	GDR	-.275	.096	-.143	-2.854	.005
	AGIST	-.212	.088	-.118	-2.402	.017

Dependent Variable: Prospects of eLearning in HEIs of KPK.

Table 7 : Showing the Excluded Variables from FIVE Models

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
5	CTY	.088(e)	1.787	.075	.095	.963
	ICTQ	-.017(e)	-.198	.843	-.011	.310
	PPR	.038(e)	.451	.652	.024	.333

d) Analysis II

The first model (table 5) explains 11% of the variation in dependent variable however, this prediction power increases gradually with the succeeding models of regression and finally reaching the level of 19% prediction of the prospects. The fifth model includes five factors as the best fit variables explaining maximum of variation in the dependent variable. The excluded variables (table 7) appear with p-values (.075, .843, and .652) which are far greater than the required threshold of .05.

The best fit is:

$PRS = a+\beta_{1CNC}+\beta_{7SNS}+\beta_{5RTPE}+\beta_{2GDR}+\beta_{8AGST}+e$

$PRS = 6.541+-.940+.446+-.322+-.275+-.212+.81488$

V. Final Analysis

Table 8 : Showing the Summary of Best-Fit Models and the Excluded Variables

OPPORTUNITES OF E-LEARNING		
1	Hypothesized Model	$OPR = a+\beta_{1CNC}+\beta_{2GDR}+\beta_{3ICTQ}+\beta_{4PPR}+\beta_{5RTPE}+\beta_{6CTY}+\beta_{7SNS}+\beta_{8AGST}+e$
2	Best Fit	$OPR = a+\beta_{1CNC}+\beta_{5RTPE}+\beta_{6CTY}+\beta_{7SNS}+e$ $OPR = 5.820+-.750+-.364+.216+.228+.71002$
3	Excluded Variables	ICTQ, PPR, GDR & AGIST
PROSPECTS OF E-LEARNING		
1	Hypothesized Model	$PRS = a+\beta_{1CNC}+\beta_{2GDR}+\beta_{3ICTQ}+\beta_{4PPR}+\beta_{5RTPE}+\beta_{6CTY}+\beta_{7SNS}+\beta_{8AGST}+e$
2	Best Fit	$PRS = a+\beta_{1CNC}+\beta_{7SNS}+\beta_{5RTPE}+\beta_{2GDR}+\beta_{8AGST}+e$ $PRS = 6.541+-.940+.446+-.322+-.275+-.212+.81488$
3	Excluded Variables	CTY, ICTQ & PPR

Table 9 : Analysis of the Role played by Demographics

	Factors	Reg-1 (OPR)	Reg-2 (PRS)	Role
1	CNC	√	√	2
2	SNS	√	√	2
3	ICTQ	-	-	0
4	RTPE	√	√	2
5	GDR	-	√	1
6	PPR	-	-	0
7	CTY	√	-	1
8	AGIST	-	√	1

In table 9 following findings emerge:

1. CNC, SNS & RTPE are the most significant factors which are playing roles in both the opportunities and prospects.
2. The respondents with 'formal and informal' ICT qualification and those from public and private HEIs view both the opportunities and prospects in a similar manner.
3. Similar opportunities are expressed by both the males and females but they are different about the prospects of eLearning.
4. There is difference of opportunities in big and small cities showing the difference of resources available in both the cities.
5. Likewise, respondents from older institutes expect different prospects than those from new institutions.

VI. Conclusions

Despite the researchers' conviction that eLearning has the potential to create current opportunities and thereby future prospects, it is not difficult to express several counterarguments against such overoptimistic conclusions (Hvorecky, 2005). More specifically, eLearning is either a threat or opportunity for the HEIs of the world in general and developing countries in particular. But the benefits are determined by the ability of developers and users to tame the technologies and change their context simultaneously as to create a customized and localized match between the requirements of eLearning and objectives of a particular institute, community, or state. This requires research on the nature of technologies, native context and the relationships between the two at the moment and in future (Nawaz & Kundi, 2010a).

The management of the university and eLearning-developers must understand the native context which contains powerful demographic diversities which, if not identified, can be counterproductive in implementing the digital systems in higher education. As table 9 shows, the divides between computer/non-computer, science and non-science and respondent type (teachers, students and administrators) alarmingly different from each other. All the three factors are playing parallel role in determining both the opportunities and prospects of eLearning. These differences in users' opinion must be addressed because they can either make or break the present and future of eLearning in Higher Education Institutions of Khyber Pakhtoonkhwa, Pakistan.

References Références Referencias

1. Abrami, P. C., Bernard , R. M., Wade, A., Schmid, R. F., Borokhovski, E., Tamim, R., Surkes, M. A., Lowerison, G., Zhang, D., Nicolaidou, I., Newman, S., Wozney, l., and Peretiatkowicz, A. (2006). A Review of e-Learning in Canada: A Rough Sketch of the Evidence, Gaps and Promising Directions. *Canadian Journal of Learning and Technology*, 32(3), Fall/Autumn. Retrieved May 14, 2007, from http://www.cjlt.ca/.
2. Cagiltay, N. E., Yildirim, S., and Aksu, M. (2006). Students' Preferences on Web-Based Instruction: linear or non-linear. *Journal of Educational Technology & Society*, 9 (3), 122-136. Retrieved April 10, 2007, from http://www.ask4research.info/
3. Chan, A. & Lee, M. J. W. (2007). We want to be Teachers, Not Programmers: In Pursuit of Relevance and Authenticity for Initial Teacher Education Students Studying an Information Technology Subject at an Australian University. *Electronic Journal for the Integration of Technology in Education*, 6, 79. Retrieved April 10, 2007, from http://ejite.isu.edu/Volume3No1/.
4. Drexler, W., Dawson, K., & Ferdig, RE. (2007) Collaborative Blogging as a Means to Develop Elementary Expository Writing Skills. *Electronic Journal for the Integration of Technology in Education, Vol. 6, p.140*.
5. *Education and ICT.* (2007). UNESCO. Retrieved October 10, 2007, from http://portal. unesco. org/ci/en/ev.php-URL_ID=19487& URL_DO= DO_TOPIC&URL_ SECTION=201.html.
6. Ezziane, Z. (2007) Information Technology Literacy: Implications on Teaching and Learning. *Journal of Educational Technology & Society,* 10 (3), 175-191. Retrieved April 10, 2007, from http://www. ask4research.info/.
7. Gay, G., Mahon, S., Devonish, D., Alleyne, P. & Alleyne, P. G. (2006). Perceptions of information and communication technology among undergraduate management students in Barbados. *International Journal of Education and*

Development using ICT, 2(4). Retrieved May 11, 2007, from http://ijedict.dec.uwi.edu//

8. Graff, M., Davies, J. & McNorton, M. (2001). Cognitive Style and Cross Cultural Differences in Internet Use and Computer Attitudes. *European Journal of Open, Distance and E-Learning*. Retrieved April 10, 2007, from http://www.eurodl.org/.
9. Gray, D. E., Ryan, M. & Coulon, A. (2003). The Training of Teachers and Trainers: Innovative Practices, Skills and Competencies in the use of eLearning. *European Journal of Open, Distance and E-Learning*. Retrieved April 10, 2007, from http://www.eurodl.org/.
10. Hagan, D. (2003). *Employer Satisfaction with ICT Graduates*. Retrieved May 28, 2007, from http://crpit.com/confpapers/CRPITV30Hagan.pdf.
11. Hvorecký, J., Manažmentu, V. S. & Cesta, P. (2005). Can E-learning break the Digital Divide? *European Journal of Open, Distance and E-Learning*. Retrieved April 10, 2007, from http://www. eurodl.org/.
12. Kanuka, H. (2007). Instructional Design and eLearning: A Discussion of Pedagogical Content Knowledge as a Missing Construct. *E-Journal of Instructional Science and Technology (e-JIST)*. 9(2). Retrieved July 18, 2007, from http://www.usq. edu.au/electpub/e-jist/docs/vol9_no2/default.htm.
13. Koo, A. C. (2008). Factors affecting teachers' perceived readiness for online collaborative learning: A case study in Malaysia. *Journal of Educational Technology & Society*, 11 (1), 266-278. Retrieved November 10, 2008, from http://www.ask4research.info/
14. Kundi, G.M. & Nawaz, A. (2010). From objectivism to social constructivism: The impacts of information and communication technologies (ICTs) on higher education. *Journal of Science and Technology Education Research. 1(2):30-36.* Available online http://www.academicjournals.org/JSTER
15. Loing, B. (2005). ICT and Higher Education. *General delegate of ICDE at UNESCO. 9th UNESCO/NGO Collective Consultation on Higher Education (6-8 April 2005)*. Retrieved June 24, 2007, from http://ong-comite-liaison.unesco.org/ongpho/acti/3/11/rendu/20/pdfen.pdf.
16. Luck, P. & Norton, B. (2005). Problem Based Management Learning-Better Online? *European Journal of Open, Distance and E-Learning*. Retrieved April 10, 2007, from http://www. eurodl.org/.
17. Lynch, J., Sheard, J., Carbone, A. & Collins, F. (2005). Individual and Organizational Factors Influencing Academics' Decisions to Pursue the Scholarship of Teaching ICT. *Journal of Information Technology Education*, 4. Retrieved July 14, 2007, from http://jite.org/documents/Vol4/.
18. Manochehr, N. (2007). The Influence of Learning Styles on Learners in E-Learning Environments: An Empirical Study. *Computers in Higher Education and Economics Review*, 18. Retrieved April 10, 2007, from http://www.economicsnetwork. ac.uk/cheer.htm.
19. Mehra, P. & Mital, M. (2007). Integrating technology into the teaching-learning transaction: Pedagogical and technological perceptions of management faculty. *International Journal of Education and Development using ICT*, 3(1). Retrieved October 11, 2007, from http://ijedict.dec.uwi.edu//.
20. Moolman, H. B., & Blignaut, S. (2008). Get set! e-Ready, … e-Learn! The e-Readiness of Warehouse Workers. *Journal of Educationa. Technology & Society*, 11 (1), 168-182. http://www.ask4research.info/ accessed April 10, 2007.
21. Nawaz, A. & G M Kundi (2010) Demographic implications for the eLearning user perceptions in HEIs of NWFP, Pakistan. *Electronic Journal of Information Systems for Developing Countries, 41(5), 1-17.* (a)
22. Nawaz, A. & Kundi, G.M. (2010) Predictor of e-learning development and use practices in higher education institutions (HEIs) of NWFP, Pakistan. *Journal of Science and Technology Education Research. 1(3):44-54.* Available online http://www.academicjournals.org/JSTER (b)
23. Nawaz, A., Kundi, G.M. & Shah, B. (2007). Metaphorical interpretations of information systems failure. *Peshawar University Teachers' Association Journal*, 14, 15-26.
24. Oliver, R. (2002). *The role of ICT in higher education for the 21st century: ICT as a change agent for education*. Retrieved April 14, 2007 from http://elrond.scam.ecu.edu.au/oliver/2002/he21.pdf.
25. Pfeffer, T. (2004). Open sources for higher education. Do information technologies change the definition of public and private goods? *CHER 17th Annual Conference, Enschede, the Netherlands, 17-19 September 2004*. Retrieved April 7, 2007 from http://www.iff.ac.at/hofo/pfeffer/2004_Pfeffer _open_sources_CHER.docQureshi et al., 2009).
26. *Policies and trends in higher education.* (2001). The University of Queensland Australia (UQA). Retrieved April 10, 2007, from http://www.tedi.uq.edu.au/largeclasses/pdfs/LitReview_6_ policies&trend.pdf
27. Sife, A. S., Lwoga, E. T. & Sanga, C. (2007). New technologies for teaching and learning: Challenges for higher learning institutions in developing countries. *International Journal of Education and Development using ICT,* 3(1). Retrieved July 21, from http://ijedict.dec.uwi.edu//.
28. Thompson, J. (2007). Is Education 1.0 Ready for Web 2.0 Students? *Innovate Journal of Online Education*, 3(4), April/May. Retrieved October 22, 2007, from http://Innovateonline.info.

29. Tinio, V. L. (2002). ICT in education. *Presented by UNDP for the benefit of participants to the World Summit on the Information Society. UNDP's regional project, the Asia-Pacific Development Information Program (APDIP), in association with the secretariat of the Association of Southeast Asian Nations (ASEAN).* Retrieved July 14, 2007 from http://www.apdip.net/publications/iespprimers/eprimer-edu.pdf.
30. Valcke, M. (2004). ICT in higher education: An uncomfortable zone for institutes and their policies. In R. Atkinson, C. McBeath, D. Jonas-Dwyer & R. Phillips (Eds), beyond the comfort zone: *Proceedings of the 21st ASCILITE Conference (pp. 20-35). Perth, 5-8 December.* Retrieved April 10, 2007, from http://www.ascilite.org.au/ conferences/perth04/procs/valcke-keynote.html.
31. Wijekumar, K. (2005). Creating Effective Web-Based Learning Environments: Relevant Research and Practice. *Innovate Journal of Online Education*, 1(5), June/July. Retrieved April 10, 2007, from http://Innovateonline.info.
32. Wims, P. & Lawler, M. (2007). Investing in ICTs in educational institutions in developing countries: An evaluation of their impact in Kenya. *International Journal of Education and Development using ICT*, 3(1). Retrieved July 21, 2007, from http://ijedict.dec.uwi.edu//.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Software & Data Engineering
Volume 12 Issue 15 Version 1.0 Year 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Simulator for Resource Optimization of Job Scheduling in a Grid Framework

By P.K. Suri & Sumit Mittal

M.M. University, Mullana, Ambala, Haryana, India

Abstract - Traditionally, computer software's has been written for serial computation. This software is to be run on a single computer with a single Central Processing Unit (CPU). A problem is broken into a discrete serial of instructions that executed in the exact order, one after another. Only one instruction can be executed at any moment of time on a single CPU. Parallel computing, on the other hand, is the simultaneous use of multiple computer resources to solve a computational problem. The program is to be run using multiple CPU's. A problem is broken into discrete parts that can be solved concurrently and executed simultaneously on different CPU's. The purpose of this proposed work is to develop a simulator using Java for the implementation of Job scheduling and shows that Parallel Execution is efficient with respect to serial execution in terms of time, speed and resources.

Keywords : *grid computing, grid framework, job scheduling, parallel computing, resource optimization.*

GJCST-C Classification : *D.4.1*

Strictly as per the compliance and regulations of:

Simulator for Resource Optimization of Job Scheduling in a Grid Framework

P.K. Suri [α] & Sumit Mittal [σ]

Abstract - Traditionally, computer software's has been written for serial computation. This software is to be run on a single computer with a single Central Processing Unit (CPU). A problem is broken into a discrete serial of instructions that executed in the exact order, one after another. Only one instruction can be executed at any moment of time on a single CPU. Parallel computing, on the other hand, is the simultaneous use of multiple computer resources to solve a computational problem. The program is to be run using multiple CPU's. A problem is broken into discrete parts that can be solved concurrently and executed simultaneously on different CPU's. The purpose of this proposed work is to develop a simulator using Java for the implementation of Job scheduling and shows that Parallel Execution is efficient with respect to serial execution in terms of time, speed and resources.

Keywords : grid computing, grid framework, job scheduling, parallel computing, resource optimization.

I. Introduction

Among the many disciplines of computer science, parallel processing is a discipline that deals with system structure and software processes related to the contingency performance of computer programs. It has been an area of active research interest and application for many fields, mainly the focus on powerful processing, but is now growing as the frequent processing model due to the semiconductor industry's move to multi-core processor chips.

Typically, software has been programmed for sequential computation, i.e., to be run on just one computer having just one main processing unit; where Instructions are implemented one after another, a problem is divided into distinct sequence of guidelines and only one instruction is executed at any instant. Although simple and economical serial computing is far much slower as compared to parallel computing. In essence, parallel computing is the simultaneous use of more than one processor or computer to solve a problem. Problems are run on multiple processors where each problem is broken into discrete parts which are further broken down into series of instructions to be executed simultaneously on different processors.

In the recent days, parallel computing has become popular based on multi-core processor chips. Most desktop computer and laptop computer systems are now delivered with dual-core micro-processors with quad-core processor chips which becomes easily available. Processor producers have started to increase overall computing performance by including additional CPU cores to provide the maximum parallelism in a program. The reason is that improving performance through parallel computing can be far more energy-efficient than improving micro-processor time wavelengths. In a world which is progressively mobile and energy aware, this has become essential.

II. Parallel Computing

Parallel computing is an outline of computation in which many data operations functions are carried out simultaneously [1].

Parallel computing takes four different forms: The first one is data parallelism / loop-level parallelism. In this computational form, a single thread controls all the data operations and in other situations, multiple threads control the execution, but they execute the same code. Second is Instruction level parallelism where instructions can be re-ordered and then, they are combined into groups and executed in parallel without affecting the result of the program [2]. The third form is task parallelism which is in contrast with data parallelism where the processor executes different threads in same or different sets of data. It focuses on allocating the processes on different parallel computing nodes. Task parallelism does not generally changes with the size of a problem [2]. The last form of parallel computation is bit level parallelism in which processors have to execute an operation on variables whose sizes are larger than the size of word.

Parallel computing has some advantages that make it attractive for certain types of problems that are suitable for use of multiprocessors, especially given limited computer memory, provides concurrency, saves money - Parallel computing resources can be built from cheap commodity components as shown in the figure 1, uses resources from a wide area network and saving time - Allocating more resources for a task shortens, it's time for completion with potential cost savings.

Conversely, parallel programming has also some disadvantages. By increasing the processors, memory in parallel computers, hence, produces a lot of data (I/O) and require parallel file system, need more space and more power, which leads to load imbalance.

Author α : HCTM Technical Campus Kaithal, Haryana, 136 027, India. E-mail : pksuritf25@yahoo.com
Author σ : M.M. Institute of Computer Technology & Business Management, M.M. University, Mullana, Ambala, Haryana, 133 207 India. E-mail : sumit_amb@yahoo.com

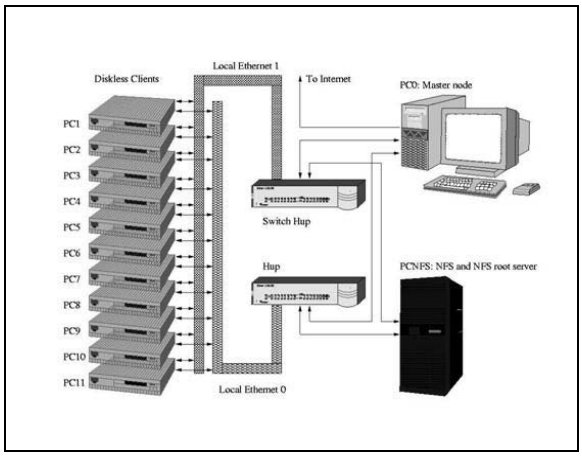


Figure 1 : Parallel Computing Cluster [3]

Parallel computer programs are more difficult to write than sequential ones, [7] because concurrency introduces several new classes of potential software bugs, of which race conditions are the most common. Communication and synchronization between the different subtasks are typically some of the greatest obstacles to getting good parallel program performance.

III. Association With Grid Framework

Grid computing combines computers from multiple administrative domains to solve a single task [4]. The grid can be thought of as a distributed system with non-interactive workloads that involve a large number of files. Grid computing tends to be more loosely coupled, heterogeneous and geographically dispersed [4]. On a single-processing machine, testing one model can take as long as five days. Using grid framework, the model can be distributed into different number of processing segments, each of which goes to its own processor; a task that normally takes five days can be completed in several hours.

IV. Comparison of Grids and Conventional Supercomputers

Distributed or grid computing in general is a special type of parallel computing that relies on complete computers resources (with onboard CPU's, storage, power supplies, network interfaces, etc.) connected to a network (private, public or the Internet) by a conventional network interface, such as Ethernet. This is in contrast to the traditional notion of a supercomputer, which has many processors connected by a local high-speed computer bus [6]. The size of a grid may vary from small network of workstations within a corporation to large public collaborations across many companies and networks. The notion of a confined grid may also be known as intra-nodes cooperation whilst the notion of a larger, wider grid may thus, refer to inter-nodes cooperation [7].

The primary advantage of distributed computing is that each node can be purchased as commodity hardware, which, when combined, can produce a similar computing resource as multiprocessor supercomputer, but at a lower cost. The primary performance disadvantage is that the various processors and local storage areas do not have high-speed connections. This arrangement is thus well-suited to applications in which multiple parallel computations can take place independently, without the need to communicate intermediate results between processors [8].

V. Simulation of Resource Optimization

The purpose of this research work is to develop an algorithm for the resource optimization of job scheduling in a grid framework and shows that parallel execution is efficient in terms of time, speed and throughput. In addition to total time taken to execute all jobs, we will also calculate the CPU's usage efficiency using the following equation:

$$Efficiency = \frac{\sum_{i=1}^{number\ \ of\ CPUs} Worktime\ of\ CPU_i}{Tolal\ number\ of\ Jobs\ \times\ Time\ for\ 1\ Job} \times 100\%$$

The proposed application consists of 3 different classes:

1. Main Class: It contains a method that is being executed at program startup. The main flow occurs in this method.
2. Processor Class: This class contains methods to assign a job, check whether processor is free and ready for the next job and to calculate the amount of time it was busy.
3. Scheduler Class: It assigns jobs to the processors using the Round Robin scheduling. When a job needs to be assigned to the CPU, scheduler searches for the free processor and assigns the job to it. If at some moment, all CPUs are busy, it waits until one of the processors becomes free.

VI. Algorithm to Compute the Efficiency of Job Scheduling in Grid Framework

Step 1. Read one line of data from the input file:

i. Test serial number
ii. Number of jobs
iii. Number of processors.

Step 2. Create an array of instances of the Processor class. Create Job scheduler instance.

Step 3. Run Job Scheduler.

Step 4. When work is finished (all jobs have been executed); and calculate the results (time taken).

Step 5. Write results to the output file.

Step 6. Go to step no. 1) unless input file is fully processed.

VII. Simulation Results and Discussions

In order to compare serial and parallel implementations, a few series of test has been performed. For a fixed number of CPU's (2, 5, 10, 20, 30, 40, 50 in different series of tests), we ran 200 tests with different number of Jobs (5, 10, 15, etc up to 1000).

The input file is a simple text file that can be created in any text editor as shown in figure 1. Each line of this file should consist of 3 values separated by pipes ("|"). First value is a string that is a serial number of the test, second value is an integer that is number of jobs and third value is an integer that is number of processors.

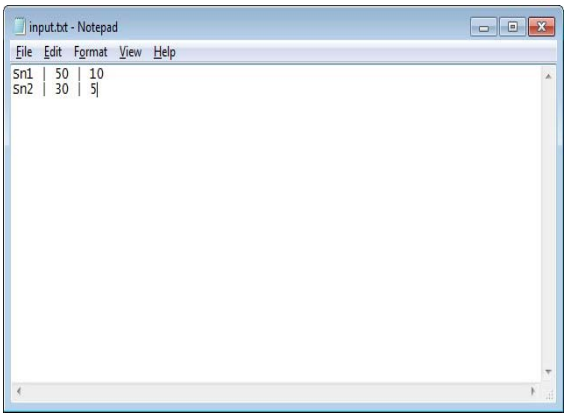


Figure 1 : Input File

Figure 2 shows the output file of the developed simulator, in this, each row consists of 5 columns separated by pipe char ("|"):

Sr. N. | Number of Input Jobs | Number of Processors | Time Taken | Execution Type (Serial / Parallel)

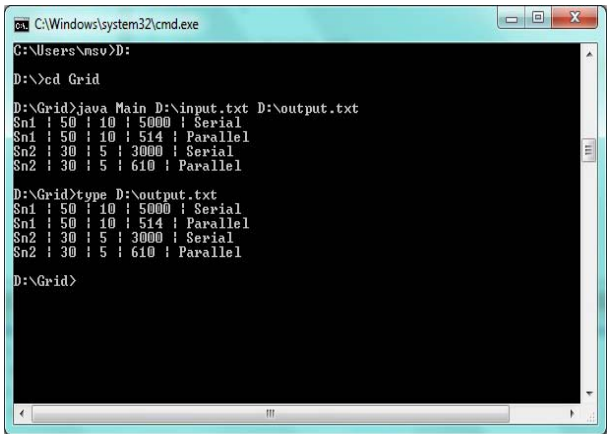


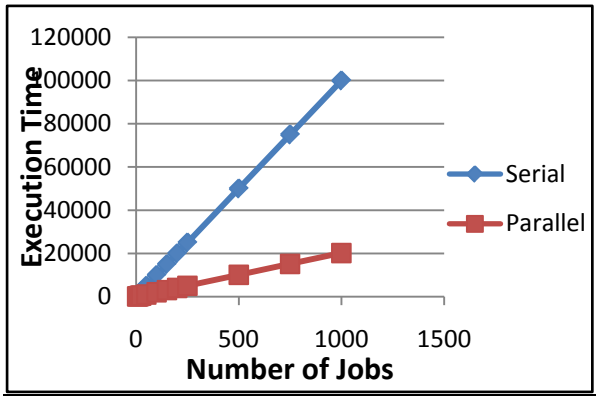
Figure 2 : Output File

Test Case 1: Table 1 shows the execution time (in microseconds) taken by the processors in serial and parallel computation and efficiency of the system for the 5 CPU's.

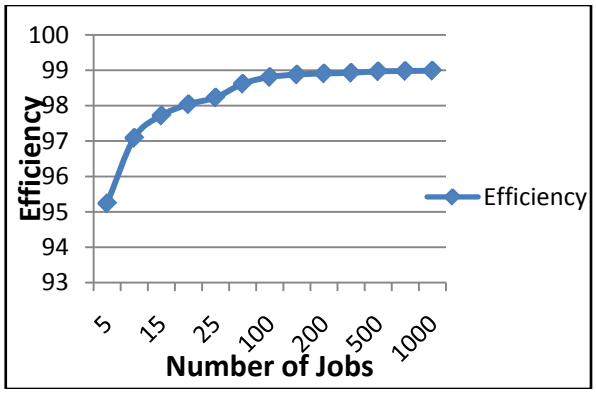
Jobs	Time (In Microseconds)		Efficiency
	Serial	Parallel	
5	500	105	95.24
10	1000	206	97.09
15	1500	307	97.72
20	2000	408	98.04
25	2500	509	98.23
50	5000	1014	98.62
100	10000	2024	98.81
150	15000	3034	98.88
200	20000	4044	98.91
250	25000	5054	98.93
500	50000	10104	98.97
750	75000	15154	98.98
1000	100000	20204	98.99

Table 1 : Execution Time and Efficiency (Number of CPU = 5)

The graph 1 depicts the relationship between the no. of jobs & the execution time estimation for the test case 1 (number of CPU's = 5) and graph 2 shows the efficiency vs. no. of jobs in terms of the time & resources.



Graph 1



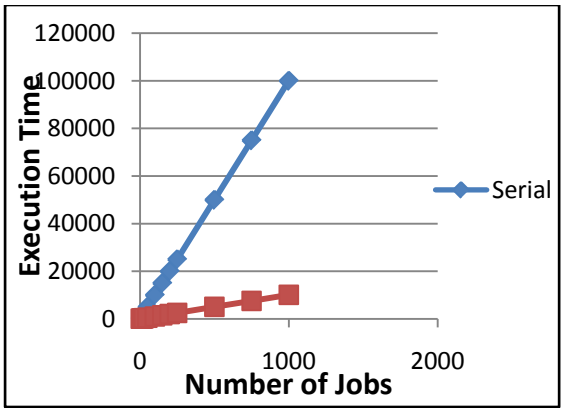
Graph No. 2

Test Case 2: Table 2 shows the execution time taken by the processors in serial and parallel computation and efficiency of the system for 10 CPU's.

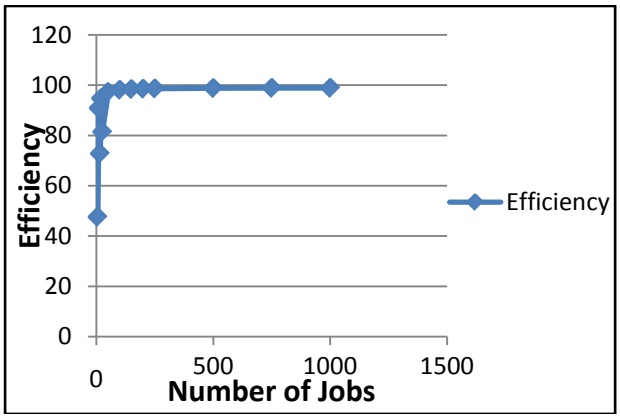
Jobs	Time (In Microseconds)		Efficiency
	Serial	Parallel	
5	500	105	47.62
10	1000	110	90.91
15	1500	206	72.82
20	2000	211	94.79
25	2500	307	81.43
50	5000	514	97.28
100	10000	1019	98.14
150	15000	1524	98.43
200	20000	2029	98.57
250	25000	2534	98.66
500	50000	5059	98.83
750	75000	7584	98.89
1000	100000	10109	98.92

Table 2 : Execution Time and Efficiency (Number of CPU = 10)

The graph 3 depicts the relationship between the number of jobs and the execution time estimation for the test case 2 (number of CPU's = 10) and graph 4 shows the efficiency vs. number of jobs in terms of the time and resources.



Graph No. 3



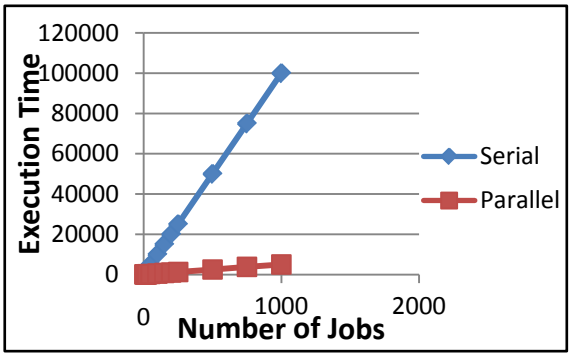
Graph No. 4

Test Case 3: Table 3 shows the execution time taken by the processors in serial and parallel computation and efficiency of the system for 20 CPU's.

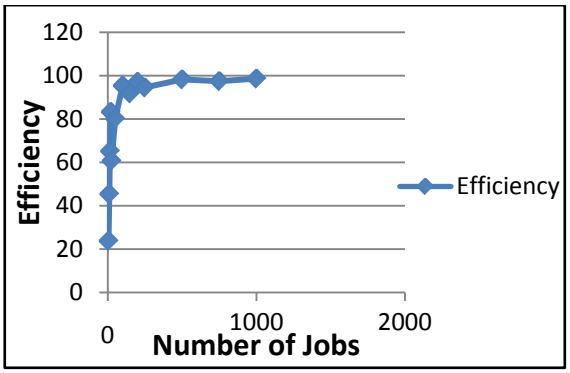
Jobs	Time (In Microseconds)		Efficiency
	Serial	Parallel	
5	500	105	23.81
10	1000	110	45.45
15	1500	115	65.22
20	2000	120	83.33
25	2500	206	60.68
50	5000	312	80.13
100	10000	524	95.42
150	15000	817	91.80
200	20000	1029	97.18
250	25000	1322	94.55
500	50000	2544	98.27
750	75000	3847	97.48
1000	100000	5069	98.64

Table 3 : Execution Time and Efficiency (Number of CPU = 20)

The graph 5 depicts the relationship between the number of jobs and the execution time estimation for the test case 3 (number of CPU's = 20) and graph 6 shows the efficiency vs. number of jobs in terms of the time and resources.



Graph No. 5

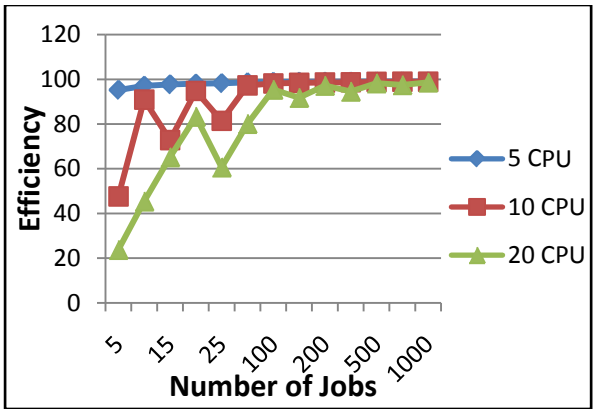


Graph No. 6

Table 4 shows the efficiency for different number of CPU's for a given set of jobs executed in the grid framework.

Jobs	Efficiency		
	5 CPU	10 CPU	20 CPU
5	95.24	47.62	23.81
10	97.09	90.91	45.45
15	97.72	72.82	65.22
20	98.04	94.79	83.33
25	98.23	81.43	60.68
50	98.62	97.28	80.13
100	98.81	98.14	95.42
150	98.88	98.43	91.8
200	98.91	98.57	97.18
250	98.93	98.66	94.55
500	98.97	98.83	98.27
750	98.98	98.89	97.48
1000	98.99	98.92	98.64

Table 4 : Comparison in terms of efficiency for different CPU's



Graph 7 : Comparison of efficiency for the different test cases

VIII. Discussion and Conclusion

After analyzing the generated results, it has been concluded that parallel execution is very efficient in terms of execution time and resources. Parallel execution on N processors is almost N times faster than serial execution. It's not exactly N times faster because of the job scheduler that needs some time to delegate tasks to the processors available. For large number of tasks, it approaches to 98% and 2% is that time when CPU's are unused waiting for a task from the job scheduler.

For a fixed number of jobs, when we are increasing the number of processors, the efficiency will be decreasing as depicts by graph no. 7, because when all processors are busy all the time; efficiency will be equal to 100%. When some of the processors were free (without a job assigned) some of the time, efficiency will be less than 100%.

As multi-core processors bring parallel computing to mainstream customers, the key challenge in computing today is to transition the software industry to parallel programming. Future capabilities such as photorealistic graphics, computational perception and machine learning rely heavily on highly parallel algorithms. Enabling these capabilities will advance a new generation of experiences that will expand the scope and efficiency of what users can accomplish in their digital lifestyles and work place. These experiences include more natural, immersive and increasingly multi-sensory interactions that offer multi-dimensional richness and context awareness.

References Références Referencias

1. Gottlieb. A, Almasi. G. S, "Highly parallel computing", Benjamin- Cummings Publishing Co., USA ©1989, ISBN:0-8053-0177-1.
2. David Culler, J.P, "Parallel computer architecture, Morgan Kaufmann Publishing Inc., San Fancisco, p. 15. ISBN 1-55860-343-3, 1997.

3. URL: http://proj1.sinica.edu.tw/~ stat phys/ computer/buildPara.html.
4. http://www.e-sciencecity.org/EN/ gridcafe/ what -is-the-grid.html
5. Pervasive and Artificial Intelligence Group publications, Diuf.unifr.ch, 2010.
6. Berman, F., Anthony J. G. H. and Geoffrey C. F, "Grid Computing: Making the global infrastructure a reality", ISBN 0-12-742503-9, 2003.
7. Asanovic, Krste et al., "The Landscape of Parallel Computing Research: A View from Berkeley", University of California, Berkeley, Technical Report No. UCB/EECS-2006-183, 2006.
8. URL: http://www.e-sciencecity.org/ EN/ gridcafe/ computational-problems.html.
9. S.V. Adve et al., "Parallel Computing Research at Illinois: The UPCRC Agenda", Parallel@Illinois, University of Illinois, Urbana, 2008.
10. Barney, Blaise. "Introduction to Parallel Computing". Lawrence Livermore National Laboratory, 2007.
11. Stallings, William (2004). Operating Systems Internals and Design Principles (fifth international edition). Prentice Hall. ISBN 0-13-147954-7.
12. Hennessy, John L.; Patterson, David A., Larus, James R. (1999). Computer organization and design : the hardware/software interface (2. ed., 3rd print. ed.). San Francisco: Kaufmann. ISBN 1-55860-428-6.
13. Hennessy, John L.; Patterson, David A. (2002). Computer architecture / a quantitative approach. (3rd ed.). San Francisco, Calif.: International Thomson. p. 43. ISBN 1-55860-724-2.
14. Bernstein, A. J., "Analysis of Programs for Parallel Processing", IEEE Transactions on Electronic Computers EC-15 (5): 757– 763, DOI: 10.1109/ PGEC. 1966.264565, 1966.
15. Roosta, Seyed H, "Parallel processing and parallel algorithms: theory and computation, New York, Springer, pp no. 114, ISBN 0-387-98716-9, 2000.

Global Journal of Computer Science and Technology
Software & Data Engineering
Volume 12 Issue 15 Version 1.0 Year 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Software Engineering: Factors Affect on Requirement Prioritization

By Shams ul Hassan & Salman Afsar Awan

University of Agriculture Faisalabad, Pakistan

Abstract - Software engineering research is yet in its early stages hence it needs evaluation. So, software engineers think about experimental research and try to adopt analytical approaches to validate results like in other sciences. It should be asserting that requirement engineering process is to use requirements prioritization. The use of requirements prioritization helps the anatomy of requirements and isolates the most important requirements. A lot of prioritization techniques, practices and methodologies are used in software requirements. But lack of empirical search program and proficient methodology, was not decide which should be implemented. In this research, the requirement prioritization for systematical reviews was carried out. Based on systematic review, a framework is introduced for further research within requirement prioritization. This paper described a framework for scrutinize the discussion that take place during requirements elicitation and requirements prioritization. The survey presented in the paper gives a practical view how to prioritize the requirements. It also reflects the requirements prioritization in the industries needs. Which factors of the requirements engineering affect the requirements prioritization.

Keywords : *framework, requirements prioritization, software engineering, requirement engineering, systematic review.*

GJCST-C Classification : *D.2.9*

Strictly as per the compliance and regulations of:

Software Engineering: Factors Affect on Requirement Prioritization

Shams ul Hassan [α] & Salman Afsar Awan [σ]

Abstract - Software engineering research is yet in its early stages hence it needs evaluation. So, software engineers think about experimental research and try to adopt analytical approaches to validate results like in other sciences. It should be asserting that requirement engineering process is to use requirements prioritization. The use of requirements prioritization helps the anatomy of requirements and isolates the most important requirements. A lot of prioritization techniques, practices and methodologies are used in software requirements. But lack of empirical search program and proficient methodology, was not decide which should be implemented. In this research, the requirement prioritization for systematical reviews was carried out. Based on systematic review, a framework is introduced for further research within requirement prioritization.

This paper described a framework for scrutinize the discussion that take place during requirements elicitation and requirements prioritization. The survey presented in the paper gives a practical view how to prioritize the requirements. It also reflects the requirements prioritization in the industries needs. Which factors of the requirements engineering affect the requirements prioritization.

Keywords : framework, requirements prioritization, software engineering, requirement engineering, systematic review.

I. Introduction

The term requirement may be defined as demand or need. In the world of the software engineering, a requirement is a explanation of what the purposed system should do or perform. A system may have a lot of requirements. Software requirements demand what must be accomplished, shaped or provided. Requirement elicitation is all about knowing the desires of stakeholders. [1] The term requirement has been used in the software engineering society since 1960. The requirements provide a firm basis for the success of the project and delivery of the product. The requirements often shrink the gap between software team and end users. Requirement phase begin at the analysis phase.

Requirements managed throughout the project life cycle. So requirements are the report of the services that a system must perform and operate under some constraints.

Author α σ : University of Agriculture (Computer Science Department), Faisalabad, 38000, Pakistan. E-mails : shams219gb@hotmail.com, salmanafsar@hotmail.com

Requirement Engineering (RE) is worried about the naming of the goals, achieved by the imagine system. Most difficult and critical to be achieve the better quality of the requirement. Incomplete, inconsistent and ambiguous requirements have the most serious impact on the required software. If requirement errors correction perform late the cost raise up to 200 times as compared the requirement errors correction perform in time. Requirement Engineering deals with a wide range of business domains and tasks like decision, administrative support. [2]

Even though the considerable Requirement Engineering (RE) explore and research attempts over the many past years but the gap between the industry and research still hang about constantly. Now Requirement Engineering research society tries to address these issues. [3] Requirement Engineering (RE) comparatively new field. Requirement Engineering is a system and processes that covers the activities based on computer system. [4] Requirement elicitation and requirement management have healthy documented using UML (Unified Modeling Language). Many tools are available that support the UML standards in any way. These tools have their own advantages and disadvantages. Many available tools need customization to meet the special requirements. The prototypes also used to create system requirements automatically. [5]

Requirement elicitation is a technique to collect the requirements. Professionals like system engineers, software analysts work with the stakeholders; this is useful for finding and solving the problems. There are many requirement elicitation techniques like interviews; questionnaire etc. [6] Requirement elicitation is the main movement in the requirement engineering process. It occupied to find out the needs and collecting the required software requirements from the stakeholders. There are some problems occurred by software engineers when they perform requirement elicitation processes. Requirement elicitation faced many problems like users' involvement and perfect documentation. To get the correct requirements and complete requirements required the right stakeholders. So there must be need to adopt the technique that could help in recognize and decide the stakeholders. [7]

It is the major problem with software developers that developed software does not meet stakeholders' requirements. It stresses the user to focus importance of

the requirements concerning the implementation cost. Time pressure and budget constraint force the software engineers to make plan for consecutive releases of software. [8] Prioritization is a method where person or group of persons rank the item in order based on their recognize importance. Prioritization is the process of starting procedure based upon urgent and long-standing need of the organizations. Prioritization leads the organization sensible plan and performs on leaders dream for the future business plan. In any good organization, it is significant to develop priorities based on IT project to move right directions. Due to time constraint and lack of budget, it is very complicated to execute all requirements that elicited during the analysis. So prioritization helps to determine which would apply first.

The objective of the current article is to knowing the factors affect on the requirement prioritization. The current survey provides the support of requirement prioritization. These requirement prioritization factors (RPF) improve the ability of the prioritization. These requirement prioritization factors (RPF) effect the cost of the imagine system.

II. Requirement Prioritization and Stakeholders

The role of prioritization of requirements is imperative to an efficient and result oriented product development. Requirement prioritization marks high risk and most important requirements to be given priority in implementation. [9] Usually stakeholder expectations are high but shortage of time, limited resources and budget constraints make it difficult to implement all requirements that have been elicited for the system. With the help of prioritization, it can be decide which one should be implement first.

In several known customers prioritization adopt difficult shapes because of different users involve as well as have different thinking and separate preferences. A most important issue arises when stakeholders are scattered different geographical areas. There priorities are not same, all the time. Every requirement analyst performs the process of prioritization. Software engineers are not well trained to elicit, gather, analyze and security requirements. [10] Requirement prioritization is extremely dangerous are of requirement engineering. Without appropriate requirement prioritization, offer by different stakeholders, the necessary objectives of the end product cannot be achieved properly. The product may fails to meet its heart objectives on the basis of several requirements prioritization techniques presented by different researchers. [11]

III. Requirement Prioritization and Agile Software Development

Agile development techniques become more accepted during last decade. Many methods built for the faster delivery of the software and those techniques ensure the developed software meets the user requirements. Requirement engineering depends on the documentation for the customer needs and agile technique depends on face to face association between the stakeholders to get the same requirements.[12] Agile software development scenarios and stories are used. Use case modeling also popular method for requirements gathering and analysis. To collect the requirements through these methods always are complete, clear and validate. [13]

In habitual software development techniques customers or stakeholders predefined their software requirements and software analysts' analysis these requirements for specification. It is very difficult and not cost effective for complete requirements. This difficulty solves by the XP methods. [14] Agile methods give the importance of continuous requirements prioritization from customer point of view. A agile approach give the client's critical role in making decision. [15]

IV. Research Method and Data Collection

The experience was drawn from this study conducted on seventeen different industries. The purpose of the survey based research was to find out how the software engineers could produce the product that might give good satisfaction of the stakeholder needs by the requirement prioritization. The organizations have different type of application domain. Through the questionnaire survey, studied the actual requirements prioritization work in special stage of software development. To get the clarification of the current requirement prioritization methods in-depth interviews were conducted. There are three groups according to their application domain.

Table 1 : Type of Application Domain

Company	Number of Employees	Application Domain
A	35	Processing
B	73	Non Processing
C	2658	IT professional

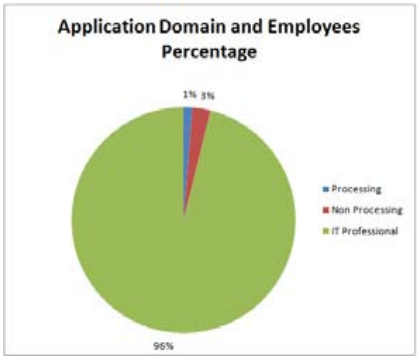


Figure 1 : Domain Expert Employees

The purpose of the focus group was to determine how organizations prioritize the requirements and which development phases involved in practice. Through the survey, it was concluded that which issues effect the prioritization and from which basis the stakeholders elicit the information on which they decide the requirement prioritization. In addition, picture was drawn to minimize the cost and enhance the effectiveness of the software though the requirement prioritization. The stakeholders and their associations to the requirement prioritization have shown the following table.

Table 2 : Company and its stakeholder's designation

Company	Relation between stakeholders and prioritization
A	Manager and Asst. Manager elicit the requirement and prioritize the requirements
B	Asst. Manager and software engineer gather information and write down the document about requirement
C	Manager, Asst. Manager and software engineers elicit the requirements, prioritize requirements and implement the actual requirements

The age of the analysts affect the result of requirement prioritization as show below.

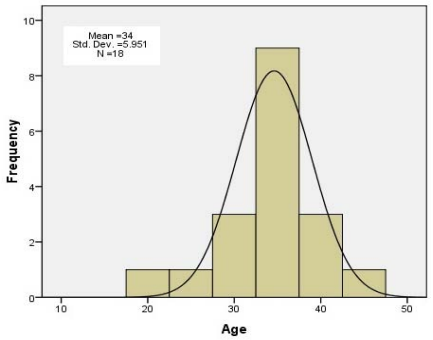


Figure 2 : Frequency Distribution

The bell-shaped histogram shows the most frequency of age counts bunches in the middle and with the counts fading off out in the tails. It is a good model for the data. The histogram helps in making decisions in which age of people is a good analysis for requirements prioritization. The most measurements lies between the 30 and 40 years age group people. The analysts who have the ages below 30 years or above 40 years did not show accurate results. Although below 30 years, analysts are so energetic and hard worker but have a little experience in the relevant field. Due to this reason they could not show cost effective results. Above 40 years analysts have so much experience but they loose their working power and activeness. Graphical presentation of the leader experiences of the participants.

Figure 3 : Leader Year Wise Experience

The leader experience plays a vital role in a good and effective software development. The leader directly attaches with customers and improves the prioritization of requirement. Leader considers the source to build the relationship to the customer. The relationship may be email, conference calls and face-to-face meeting. The leader should conduct effective meetings. Leader should plan the requirement session. He / She guides focal cause analysis and implement the requirements prioritization effort. The leader reviews the reports regularly to check the valid requirements. He / She make sure the requirements prioritization addressed by the customer according to the contract and balance with the stakeholder expectations. On the base of accurate requirements prioritization controls the

future software direction. Leader experience with the remote customer and team is decidedly required. Leadership must have the both project management as well as people management experience. Leader must have knowledge that software architecture and development based on requirements prioritization. The leader should have the ability to face the elicitation and prioritize the requirements. He / She must have the capability to identify the requirements prioritization constraints.

Qualification of the analyst is also considered main thing in solving the requirements prioritization problems.

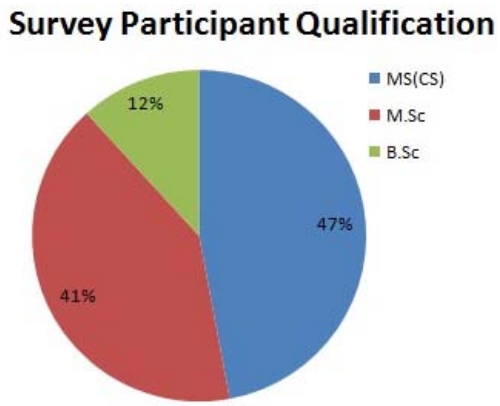


Figure 4 : Participant qualification

The team leader or business analyst required at least bachelor's degree. A survey showed that the in-house development industry try to hire at least bachelor employees for analysis, requirements analysis and prioritization but industries desired for business analysts having brilliant computer education like M.Sc. (CS) etc. Because their professional work involves solutions to meet the customer business needs and customer challenges. As industries have over all well educated and experienced persons in different working areas and industries businesses enhance day to day. In the result of this requirements of industries generate dynamically. To control and prioritized all those requirements, the team leader should be high qualified and have ability to translate industries requirements into system requirements with prioritization for software developers. They should have expertise in open range of business effect and software programs. Education generates the ability in analyst to understand the business requirements, identify those requirements, document those requirements and prioritized them for business application for software developers. They get the ability to cope the relationship between different departments, effective communication, strategy of development those prioritized requirements.

The importance of the requirements is never neglected. The requirements prioritization is based on the following factors.

Figure 5 : Purposed Frameworks

Study indicated that there are three factors which affect the priorities. It is the base line for the profit of any organization. These factors minimize the gap between the stakeholders. Any issue like relationships, communication problem and project involvement between the customers and analysts can be shrinking using these factors. There is no defined method for in-house development to prioritize the requirements. The organization must know these attributes in any software engineers. Any software developer who contains these attributes can make easily requirements prioritization and logical implementation of these requirements. The skill and education attributes cover the geographical region made their requirements priorities. It is not easy to say which individual factors affect the prioritization of the requirements.

V. Conclusion

The study has been conducted at three types of industries, to find out the results about requirements prioritization that gives the high level satisfaction of the customer. To achieve the explanation of the actual requirements prioritization, conduct the survey indepth. The purpose of the survey was to determine how organizations prioritize the requirements. The term priority is the property or attribute of the requirement. In the survey organization, the requirements elicitation and requirements prioritization interacts the domain expert users.

The survey study indicated that there are three factors like analyst's qualification, experience and age which affect the prioritization. These factors help to minimize the distance between the stakeholders. The defined three factors resolved the issues like relationships, communication problems between users and software engineers. Any software developer pursuing these attributes can make easily requirements prioritization and implementation of the requirements. Accurate requirements prioritization produced the cost effective solution for the organization.

References Références Referencias

1. Sadiq .M and Mohd .S (2009), "Elicitation and Prioritization of Software Requirements". Internation Journal of Recent Trends in Engineering, Vol.2, No.3, pp. 138-142.
2. Lamsweerde, A.V (2000), Requirement Engineering in the Year 00: A Research Perspective. Proc. 22nd Int. Conf. on the Software Engineering, University of Limerick, Ireland.
3. Sparrow .L., et. al. (2006), the Slow Wheel of Requirements Engineering Research: Responding to the Challenges of Societal Change. Deakin University, Australia.
4. Arif .S., et. Al. (2010), Requirements Engineering Processes Tools/Technology, & Methodologies. Int. Journal of Reviews in Computing. Islamabad, Pakistan.
5. Meisenbacher .l.K. (2005), The Challenges of Tool Integration for requirements Engineering. Proc. of SREO'05. France.
6. Ganesh and Gunda (2008), Requirement Engineering: Elicitation Techniques. Dept. of Technology, Mathematics and Computer Science. Sweden.
7. Rozilawati .A and Fares .A (2011), Selecting he Right Stakeholders for Requirements Elicitation: A Systematic Approach. Journal of theortical and Applied Information Technology. Malaysia.Vol.33, No.2, pp. 250-257.
8. Viggo .A (2005), An Experiment Comparision of Five Prioritization Methods Master Thesis Software Engineering, Sweden.
9. Joachim .K and Kevin .R (1997), A Cost-Value Approach for Prioritizing Requirements, IEEE Software, Germany.
10. Firesmith .D.G., (2003), Engineering Security Requirements. Journal of Object Technology. U.S.A. Vol.2, No.1, pp. 53-68.
11. Ramzan .M., et. al.(2011), Value Based Intelligent Requirements prioritization (VIRP): Expert Driven Fuzzy Logic Based Prioritization Technique. Int. Journal. Innovative Computing, Information and Control. Islamabad, Pakisatn. Vol.7, No.3, pp 1017-1038.
12. Paetsch .F., et. al. (2003), Requirement Engineering and Agile Software Development. Proc. 20th IEEE Int. Workshop on enabling technologies: Infrastructure for Collaborative Enterprises (WETICE'03). Washington, USA.
13. Firesmith .D (2004), Generating Complete. Unambiguous, and Verifiable Requirements from Stories, Scenarios and Use Cases. Journal of Object Technology. U.S.A. Vol.3, No.10, pp. 27-39.
14. Chetankumar .P and Muthu .R (2009), Story Card Based agile Software Development. Int. Journal of Hybride Information Technology. Vol.2, No.2, pp.125-140.
15. Reprioritization the Requirements in Agile Software Development: towards a conceptual model from clients perspective. Available at http://eprints. eemcs. utwente. nl/16055/01/SEKE-Rac heva-Daneva.pdf

This page is intentionally left blank

Global Journal of Computer Science and Technology
Software & Data Engineering
Volume 12 Issue 15 Version 1.0 Year 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Classification Rules and Genetic Algorithm in Data Mining

By Mr. Puneet Chadha & Dr. G.N. Singh

Panjab University, Chandigarh

Abstract - Databases today are ranging in size into the Tera Bytes. It is an information extraction activity whose goal is to discover hidden facts contained in databases. Typical applications include market segmentation, customer profiling, fraud detection, evaluation of retail promotions, and credit risk analysis. Major Data Mining Tasks and processes include Classification, Clustering, Associations, Visualization, Summarization, Deviation Detection, Estimation, and Link Analysis etc. There are different approaches and techniques used for also known as data mining models and algorithms. Data mining algorithms task is discovering knowledge from massive data sets. In this paper, we are focusing on Classification process in Data Mining.

GJCST-C Classification : *H.2.8*

Strictly as per the compliance and regulations of:

Classification Rules and Genetic Algorithm in Data Mining

Mr. Puneet Chadha [α] & Dr. G.N. Singh [σ]

I. Introduction

Databases today are ranging in size into the Tera Bytes. It is an information extraction activity whose goal is to discover hidden facts contained in databases. Typical applications include market segmentation, customer profiling, fraud detection, evaluation of retail promotions, and credit risk analysis. Major Data Mining Tasks and processes include Classification, Clustering, Associations, Visualization, Summarization, Deviation Detection, Estimation, and Link Analysis etc. There are different approaches and techniques used for also known as data mining models and algorithms. Data mining algorithms task is discovering knowledge from massive data sets. In this paper, we are focusing on Classification process in Data Mining.

The management and analysis of information and using existing data for correct prediction of state of nature for use in similar problems in the future has been an important and challenging research area for many years. Information can be analyzed in various ways. Classification of information is an important part of business decision making tasks. Many decision making tasks are instances of classification problem or can be formulated into a classification problem, viz., prediction and forecasting problems, diagnosis or pattern recognition. Classification of information can be done either by statistical method or data mining method.

II. Classification

Classification is a form of Data Analysis that can be used to construct a Model, which can be further used in future to predict the Class Label of new Datasets. Various Application of classification includes Fraud Detection, Target Marketing, Performance Prediction, Manufacturing and Medical Diagnosis.

Data Classification is a two step process

(i) The first step is a learning step. In this step a classification algorithm builds the Classifier by analyzing (or learning from)a training set made up of database tuples and their associated Class Labels. In this first step a Mapping Function Y=f(X) is learned that can predict the associated Class Label Y of a given tuple X. That mapping function or Classifier can be in the form of Classification Rules, Decision Trees or Mathematical Formulae.

(ii) Next Step of Classification, Accuracy of a Classifier is predicted. For this another set of tuples apart from training tuples are taken called as Test Sets. Then these set of tuples of test set are given as input to the Classifier.

The Accuracy of a Classifier on a given test set is the percentage of test set Tuples that are correctly classified by the Classifier.

a) Classification Methods in Data Mining

There are various Classification Methods as listed below

i. *Classification by Decision Tree Induction*

In this method Decision Tree is learned from Class-Labeled training tuples and then it is used for Classification. A Decision Tree is a flowchart–like tree structure, where each Internal Node (nonleaf node) denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node holds a class label.

While learning the Decision Tree or we can say during Tree Construction, attribute selection measures are used to select the attribute that best partitions the tuples into distinct classes. Once Decision Trees are built, Tree Pruning attempts to identify and remove branches that may reflect noise or outliers in the training data.

Learned Decision Trees can be used for Classification. Given a tuple X for which the associated Class Label is unknown, the attribute values of the tuple are tested against the decision tree. A path is traced from the root to a leaf node, which holds the class prediction for that tuple. Decision trees can be easily converted to Classification Rules.

ii. *Bayesian Classification*

Classifiers made using Bayesian Classification can predict the probability that a given tuple belongs to a particular Class.

Baye's Theorem: Using Baye's theorem we can predict Posterior Probability, P(H,X) from P(H),P(X|H) and P(X).Here X is a data tuple.

Baye's Theorem is

Author α : Assistant Professor, DAV college, Sector-10, affiliated to Panjab University, Chandigarh-160010.

Author σ : Head Department of Physics and Computer Science, Sudarshan Degree College, Lalgaon Distt. Rewa (M.P.) India.

$$P(H|X) = \frac{P(X|H)\,P(H)}{P(X)}$$

Where H ->Hypothesis such as that the data tuple X belongs to a specified Class C

P(H|X) ->Probability that hypothesis H exists for some given values of X's attribute.

P(X|H) ->Probability of X conditioned on H

P(H) ->Probability of H

P(X) ->Probability of X

Naive Bayesian Classification

Let d be a training set of tuples and $X=(x_1,x_2,x_3.....x_n)$ are the n attributes.

Let there be m classes $C_1,C_2.....C_m$.Naive Bayesian Classifier predicts that tuple X belongs to the class Ci if and only if

$$P(Ci|X) > P(Cj|X) \text{ for } 1<=j<=m,\ j \neq i$$

i.e X belongs to the Class having the Highest Posterior Probability.

Bayesian Belief Networks

The Naive Bayesian Classifier assumes Class Conditional Independence, but in practice dependencies can exist between Attributes (variables) or the tuple ie a particular categorization can depend on the values of two attributes.

Bayesian Belief Networks specify Joint Conditional Probability Distributions.

A Belief Network is defined by two components- A Directed Acyclic Graph and a Set of Conditional Probability Tables.

Each Node in the Graph correspond to actual attributes given in the data or to hidden variables. Each Arc represents a Probabilistic Dependence. Each variable is only Conditionally Dependent on its Immediate Parents. A Belief Network has one Conditional Probability Table (CPT) for each variable or attribute. In this the Conditional Probability for each known value of attribute is given for each possible combination of values of its Immediate Parents.

Trained Bayesian Belief Networks can be used for Classification. Various Algorithms for learning can be applied to the Network, rather than returning a Single Class Label.

The Classification Process can return a Probability Distribution that gives the Probability of each Class.

iii. *Rule Based Classification*

Rule Based Classifiers uses a set of IF-Then Rules for Classification.

IF Condition Then Conclusion.

The IF part is the Rule Antecedent or Precondition. Then Part is the Rule Consequent.

The Condition consists of one or more Attribute Tests that are logically ANDed. The Rule's Consequent contains a Class Prediction.

Let D be the Training data set. Let X be a tuple. If a Rule is satisfied by X, the rule is said to be triggered. Rule fires by returning the Class Prediction.

Rule Extraction from a Decision Tree

To Extract Rules from a Decision Tree, One Rule is created for each path from the root to a leaf node. Each Splitting Criterion along a given path is Logically ANDed to form the Rule Antecedent(IF part).The Leaf node holds the Class Prediction, forming the Rule Consequent(Then part).

Rule Induction using a Sequential Covering Algorithm

Here the Rules are Learned Sequentially, One at a time (for one Class at a time) directly from the Training Data (i.e without having to generate a Decision Tree first) using a Sequential Covering Algorithm.

iv. *Classification by Backpropagation*

Backpropagation is the most popular Neural Network Learning Algorithm. Neural Network is a set of connected input/output units, in which each connection has a weight associated with it. During the learning phase, the Network learns by adjusting the weights so as to be able to predict the Correct Class Label of the Input Tuples. Backpropagation performs on Multilayer Feed-Forward Neural Network. Several techniques have been developed for the Extraction of Rules from Trained Neural Networks. These factors contribute toward the usefulness of Neural Networks for Classification and Prediction in Data Mining.

v. *Support Vector Machines*

Support Vector Machines is a promising new method for the Classification of both Linear and Non Linear Data. Support Vector Machine is an algorithm that uses a Non Linear Mapping to transform the original training data into a higher dimension. Within this new dimension, it searches for Linear Optimal Separating Hyperplane (that is, a decision boundary separating the tuples of one class from another).The SVM finds this Hyperplane using Support Vectors (essential training tuples) and Margins (defined by the support vectors).

vi. *Associative Classification*

In Associative Classification, Association Rules are generated and analyzed for use in Classification. The general idea is that we can search for Strong Associations between Frequent Patterns (conjunctions of attribute-value pairs) and Class Labels. Because Association Rules explore highly confident Associations among Multiple Attributes, this approach may over-come some constraints introduced by Decision-Tree Induction which considers only one attribute at a time. In

particular three main methods are studied CBA, CMAR and CPAR.

vii. *Lazy Learners*

Lazy Learners do less work when a training tuple is presented and more work when making a Classification .When given a training tuple, a Lazy Learner simply stores it(or does a little minor processing) and waits until it is given a test tuple. Only when it sees the test tuple does it perform generalization in order to classify the tuple based on its similarity to the stored training tuples. Two examples of lazy learners are K-Nearest-Neighbour Classifiers and Case-Based Reasoning Classifiers.

III. Genetic Algorithm

A genetic algorithm (GA) is a search technique used in computing to find exact or approximate solution to optimization and search problems. Genetic algorithms are categories as global search heuristics.

Genetic algorithms are a probabilistic search and evolutionary optimization approach. Genetic algorithms are inspired by Darwin's theory about evolution. Solution to a problem solved by genetic algorithms is evolved.

Algorithm is started with a set of solutions (represented by chromosomes) called population. Solutions from one population are taken and used to form a new population. This is motivated by a hope, that the new population will be better than the old one. Solutions which are selected to form new solutions (offspring) are selected according to their fitness - the more suitable they are the more chances they have to reproduce.

This is repeated until some condition (for example number of populations or improvement of the best solution) is satisfied.

1. **[Start]** Generate random population of *n* chromosomes (suitable solutions for the problem)
2. **[Fitness]** Evaluate the fitness *f(x)* of each chromosome *x* in the population
3. **[New population]** Create a new population by repeating following steps until the new population is complete
    1. **[Selection]** Select two parent chromosomes from a population according to their fitness (the better fitness, the bigger chance to be selected)
    2. **[Crossover]** With a crossover probability cross over the parents to form a new offspring (children). If no crossover was performed, offspring is an exact copy of parents.
    3. **[Mutation]** With a mutation probability mutate new offspring at each locus (position in chromosome).
    4. **[Accepting]** Place new offspring in a new population
4. **[Replace]** Use new generated population for a further run of algorithm
5. **[Test]** If the end condition is satisfied, **stop**, and return the best solution in current population
6. **[Loop]** Go to step **2**

a) Genetic Algorithms and Classification

The construction of a classifier requires some parameters for each pair of attribute value where one attribute is the class attribute and another attribute is selected by the analyst. These parameters may be used as intermediate result for constructing the classifier. Yet, the class attribute and rest all attributes that analyst considers as relevant attributes must be the attributes of the tables that might be used for analysis in future. Hence, attribute values of class attribute are always frequent. When pre-computing the frequencies of pairs of frequent attribute values, the set of computed frequencies should also include the frequencies that a potential application needs as values of the class attribute and relevant attribute are typically frequent.

A framework for Genetic Algorithm to be implemented for Classification is

1. Start
2. Initialize the Population
3. Initialize the program size
4. Define the fitness f_i of an individual program corresponds to the number of hits and is evaluated by specific formula:
5. Run a tournament to compare four programs randomly out of the population of programs
6. Compare them and pick two winners and two losers based on fitness
7. a) Copy the two winners and replace the losers
    b) With Crossover frequency, crossover the copies of the winners
    c) With Mutation frequency, mutate the one of the programs resulting from performing step 7(a)
    d) With Mutation frequency, mutate the other of the programs resulting from performing step 7(a)
8. Repeat through step 5 till termination criteria are matched.

References Références Referencias

1. Piatetsky, Gregory. (2007). “Data mining and knowledge discovery 1996 to 2005: overcoming the hype and moving from “university” to “business” and “analytics”. *Data Min Knowl Disc (2007)* 15:99–105
2. Kriegel, Hans-Peter. Borgwardt, Karsten M. Kröger, Peer.*(et. al.)*(2007).“Future trends in data mining”. *Data Min Knowl Disc (2007)* 15:87–97
3. Luo, Qi. (2008). “Advancing Knowledge Discovery and Data Mining” *Knowledge Discovery and Data Mining, 2008.* WKDD 2008.

4. Weiss, Gary M. Zadrozny, Bianca. Saar-Tsechansky, Maytal. (2008). "Guest editorial: special issue on utility-based data mining". *Data Min Knowl Disc (2008)* 17:129–135.
5. Weber, Ben G. Mateas, Michael (2009). "A Data Mining Approach to Strategy Prediction" *978-1-4244-4815 2009 IEEE.*
6. Das, Sufal. Saha, Banani(2009). "Data Quality Mining using Genetic Algorithm". *International Journal of Computer Science and Security*, (IJCSS) Volume (3): Issue (2).
7. Kamble, Atul (2010). "Incremental Clustering in Data Mining using Genetic Algorithm". *International Journal of Computer Theory and Engineering*, Vol. 2, No. 3, June, 2010.
8. Kantarcıoglu, Murat. Xi, Bowei. Clifton, Chris. (2011). "Classifier evaluation and attribute selection against active adversaries" *Data Min Knowl Disc (2011)* 22:291–335.

Global Journal of Computer Science and Technology
Software & Data Engineering
Volume 12 Issue 15 Version 1.0 Year 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Artificial System for Prediction of Student's Academic Success from Tertiary Level in Bangladesh

By Linkon Chowdhury & Shahana Yeasmin

BGC Trust University Bangladesh

Abstract - Every year a large scale of students in Bangladesh enrol in different Universities in order to pursue higher studies. With the aim to build up a prosperous career these students begin their academic phase at the University with great expectation and enthusiasm. However among all these enthusiastic and hopeful bright students many seem to become successful in their academic career and found to pursue the higher education beyond the undergraduate level. The main purpose of this research is to develop a dynamic academic success prediction model for universities, institutes and colleges. In this work, we first apply chi square test to separate factors such as gender, financial condition and dropping year to classify the successful from unsuccessful students. The main purpose of applying it is feature selection to data. Degree of freedom is used to P-value (Probability value) for best predicators of dependent variable. Then we have classify the data using the latest data mining technique Support Vector Machines(SVM).SVM helped the data set to be properly design and manipulated. After being processed data, we used the MATH LAB for depiction of resultant data into figure. After being separation of factors we have had examined by using data mining techniques Classification and Regression Tree (CART) and Bayes theorem using knowledge base. Proposition logic is used for designing knowledge base. Bayes theorem will perform the prediction by collecting the information from knowledge Base. Here we have considered most important factors to classify the successful students over unsuccessful students are gender, financial condition and dropping year. We also consider the sociodemographic variables such as age, gender, ethnicity, education, work status, and disability and study environment that may influounce persistence or academic success of students at university level. We have collected real data from Chittagong University Bangladesh from numerous students. Finally, by mining the data, the most important factors for student success and a profile of the typical successful and unsuccessful students are identified.

Keywords : *Chi Value, P-value, CART, SVM, Bayes theorem, Degree of freedom.*

GJCST-C Classification : *I.2.0*

Strictly as per the compliance and regulations of:

Artificial System for Prediction of Student's Academic Success from Tertiary Level in Bangladesh

Linkon Chowdhury [α] & Shahana Yeasmin [σ]

Abstract - Every year a large scale of students in Bangladesh enrol in different Universities in order to pursue higher studies. With the aim to build up a prosperous career these students begin their academic phase at the University with great expectation and enthusiasm. However among all these enthusiastic and hopeful bright students many seem to become successful in their academic career and found to pursue the higher education beyond the undergraduate level. The main purpose of this research is to develop a dynamic academic success prediction model for universities, institutes and colleges. In this work, we first apply chi square test to separate factors such as gender, financial condition and dropping year to classify the successful from unsuccessful students. The main purpose of applying it is feature selection to data. Degree of freedom is used to P-value (Probability value) for best predicators of dependent variable. Then we have classify the data using the latest data mining technique Support Vector Machines(SVM).SVM helped the data set to be properly design and manipulated. After being processed data, we used the MATH LAB for depiction of resultant data into figure. After being separation of factors we have had examined by using data mining techniques Classification and Regression Tree (CART) and Bayes theorem using knowledge base. Proposition logic is used for designing knowledge base. Bayes theorem will perform the prediction by collecting the information from knowledge Base. Here we have considered most important factors to classify the successful students over unsuccessful students are gender, financial condition and dropping year. We also consider the sociodemographic variables such as age, gender, ethnicity, education, work status, and disability and study environment that may in-flounce persistence or academic success of students at university level. We have collected real data from Chittagong University Bangladesh from numerous students. Finally, by mining the data, the most important factors for student success and a profile of the typical successful and unsuccessful students are identified.

Keywords : Chi Value, P-value, CART, SVM, Bayes theorem, Degree of freedom.

I. Introduction

Student retention is an indicator of academic performance and enrolment management of the university. Increasing student retention or persistence is a long term goal in all academic institution. High rates of student attrition have been a concern for the past several years at the Chittagong University Institute of Computer Science and Engineering. Level of retention rate goes higher by various reasons and one the most important reasons are government funding, scholarship in the tertiary education environment. There are both academic and administrative staff are under pressure to come up with strategies that could increase retention rates on their courses and programs. The lowest student retention rates at all institutions of higher education are first-year students, who are at greatest risk of dropping out in the first term or semester of study or not completing their programs or degree. Therefore most retention studies address the retention of first-year students. Consequently, the early identification of vulnerable students who are prone to drop their courses is crucial for the success of any retention strategy. This would allow educational institutions to undertake timely and pro-active measures. Once identified, these at risk students can be then targeted with academic and administrative support to increase their chance of staying on the course.

Data from 2004 to 2010, covering over 200 enrolled students stored in the Computer Science and Engineering was used to perform a quantitative analysis of study outcome. There were three main types of objectives conducted in this study. The first approach is descriptive which is concerned with the nature of the dataset such as the frequency table and the relationship between the attributes obtained using cross tabulation analysis (contingency tables). In addition, feature selection is conducted to determine the importance of the prediction variables for modelling study outcome. The third type of data mining approach, i.e. predictive data mining is conducted by using two different types of classification trees. The classification tree models have some advantages. Secondly, the classification tree models are non-parametric and can capture nonlinear relationships and complex interactions between predictors and dependent variable. We decided not to use other data mining techniques such as neural networks and support vector machines even though in some cases they could achieve higher accuracy, because their structure is not transparent and usually described as a black box. It is also difficult to explain their results and how they work to a user who would like

Author α : Chittagong, Bangladesh.
E-mail : linkon_cse_cu@yahoo.com
Author σ : Chittagong, Bangladesh. E-mail : bubly.csecu@gmail.com

to apply them to a new set of data. Finally, a comparison between these models was conducted to determine the best model for the dataset. The population of this study was entering CSE in the first - year from 2004 through 2009. These were full-time degree-seeking students. Entering in the first-year session were excluded from the study for a few reasons. The independent or predictor variables fell into three main categories. These were the students' personal information, high school background. The personal information recorded for each student was:

- Ethnicity(Bengali, Marma, Chakma and Others)
- Gender
- Age
- Financial Status
- Work Status
- Disability

II. Collected Statistical Data

As part of the data-understanding phase we carried out the data on the table 1 and table 2. The Table 2 reports the results. Based on the results shown majority of Information Systems students are female (over 38%). However, percentage of female students who successfully complete the course are higher (41%) which suggests that female students are more likely to[3] pass the course than their male counterpart. When it comes to age over 26% of students are above 24. This age group is also more likely to fail the course because their percentage of students who failed the course in this age group (11.9%) is higher than their overall participation in the student population (26.2%). Statistical data on 42 students:

Table 1 : Total Outcomes

Pass	29
Fail	13

Table 2 : Descriptive statistics (percentage) – Study outcome (42 students)

Variable	Domain Name	Count	Total	Pass	Fail
Gender	Male	26	61.9	58.6	69.2
	Female	16	38.1	41.4	30.8
Age Group	>24	11	26.2	20.7	11.9
	<=24	31	73.8	79.3	61.5
Disabilities	Yes	1	2.4	0.0	7.7
	No	41	97.6	100.0	92.3
Financial Support	Yes	23	54.8	62.1	38.5
	No	19	45.2	37.9	61.5

Students with it are more likely to fail than those without it. There are huge differences in percentage of students who successfully completed [4] the course depending on their ethnic origin. A substantial number of students (over 55%) have financial support more vulnerable than the other two categories in this variable.

III. Support Vector Machines

Support Vector Machine (SVM) is one of the latest clustering techniques which enables machine learning concepts to amplify predictive accuracy in the case of axiomatically diverting data those are not fit properly. It uses inference space of linear functions in a high amplitude feature space, trained with a learning algorithm. It works by finding a hyper plane that linearly separates the training points, in a way such that each resulting subspace contains only points which are very similar. First and foremost idea behind Support Vector Machines (SVMs) is that it constituted by set of similar supervised learning. An unknown tuple is labeled with the group of the points that fall in the same subspace as the tuple. Earlier SVM was used for Natural Image processing System (NIPS) but now it becomes very popular is an active part of the machine learning research around the world. It is also being used for pattern classification and regression based applications. The foundations of Support Vector Machines (SVM) have been developed by V.Vapnik.

SVM is very effective in various data and information classification process. An expert should bear in mind two important factors for implementing SVM, these two factors or techniques are mathematical programming and kernel functions. Kernel methods leads or portrayal data into colossal amplitude margins in the anticipation that in this colossal amplitude margin the data could become more easily separated or better structured. Mathematical Programming refers the conception of the Linear programming for the best fit of Hyper plane. The word programming means to plans or make a time table for regular work. Integer Linear programming (ILP) which is the part of linear programming is very useful analytical and engineering tools to get an optimal solution .The parameters are found by solving a quadratic programming problem with linear equality and inequality constraints; rather than by solving a nonconvex, unconstrained optimization problem. The flexibility of kernel functions allows the SVM to search a wide variety of hypothesis spaces. The for-most reasons of using SVM are to select the proper Support Vectors for the data classification. The figure 1 shows a graphical view of Support Vectors selection of the process. All hypothesis space help to identify the Maximum Margin Hyper plane (MMH) which enables to classify the best and almost correct data the following figure shows the process of SVMs selection from large amount of SVMs.

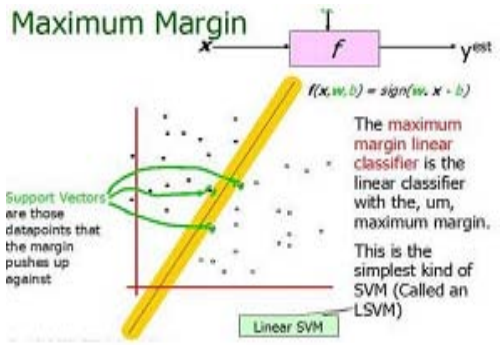


Figure 1 : Representation of Support Vectors

We can calculate the weight boundary maximum margin by using the following equation:

$$\text{margin} \equiv \arg\min_{\mathbf{x}\in D} d(\mathbf{x}) = \arg\min_{\mathbf{x}\in D} \frac{|\mathbf{x}\cdot\mathbf{w}+b|}{\sqrt{\sum_{i=1}^{d} w_i^2}}$$

Another interesting question is why maximum margin? There are some good explanations which include better empirical performance. Another reason is that even if we've made a small error in the location of the boundary this gives us least chance of causing a misclassification. The other advantage would be avoiding local minima and better classification. The goals of SVM are separating the data with hyper plane and extend this to non-linear boundaries using kernel trick [8] [11]. For calculating the SVM we see that the goal is to correctly classify all the data. For mathematical calculations we have,

[a] If $Y_i = +1$; $wx_i + b \geq 1$

[b] If $Y_i = -1$; $wx_i + b \leq 1$

[c] For all i; $y_i (w_i + b) \geq 1$

In this equation x is a vector point and w is weight and is also a vector. So to separate the data [a] should always be greater than zero. Among all possible hyper planes, SVM selects the one where the distance of hyper plane is as large as possible. If the training data is good and every test vector is located in radius r from training vector. Now if the chosen hyper plane is located at the farthest possible from the data [12]. This desired hyper plane which maximizes the margin also bisects the lines between closest points on convex hull of the two datasets. Thus we have [a], [b] & [c].

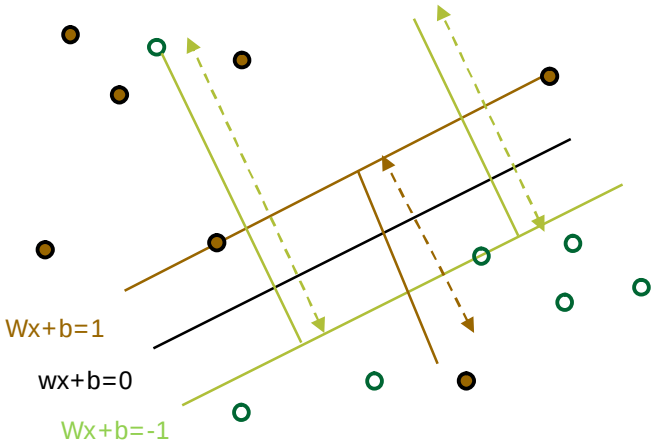


Figure 2 : Representation of Hyper planes

Distance of closest point on hyper plane to origin can be found by maximizing the x as x is on the hyper plane. Similarly for the other side points we have a similar scenario. Thus solving and subtracting the two distances we get the summed distance from the separating hyper plane to nearest points. Maximum Margin = M = 2 / ||w||

IV. Classification and Regression Tree (CART)

Classification and Regression Tree (CART) has many advantages over classification methods. It is naturally non-parametric. CART can handle [5][6] numerical data that are highly skewed. It eliminates analyst time, which would otherwise be spent determining whether variables are normally distributed. CART identifies "splitting" variables based on a complete search of all possibilities. CART is able to search all possible variable splitters, even in problems with many hundreds of possible predictors as well as dealing missing variable. Finally, CART trees are moderately simple and non-statisticians.

The purpose of an analysis based on classification tree is to find out the factors that contribute the separation of successful and unsuccessful students among the all recorded data. When the classification tree is generated it is possible to calculate the probability of each student's outcome. If once the classification tree is formed, it is possible to predict the study outcome for newly enrolled student. In each tree node the percentages for successful and unsuccessful students is given and also absolute size of the node. The variable names above the node are the predictor's criteria that make the split for the node according to classification and regression tree. Each node is split according to predictor criteria. The searching is stops when the split with the largest improvement in goodness of fit. Possible predictor variables in the dataset were included in the classification tree in splits process, which is detected in feature selection criteria. We used two stopping criteria in the training process:

1. A maximum tree depth has been proceeding into 3 levels for CHAID tree.
2. A maximum tree depth has been proceeding into 4 levels for CART tree.

Lastly for each classification tree we have assigned different costs to the classification outcomes i.e.: classification matrix. It is one of the processes of increasing the correctly classified student's outcomes. We categories the student's outcome according to Pass and Fail in the academic courses

V. Our Contribution

In this research we explore the concepts and technique of SVM to classify the data collected in our experiments. We have approximately collected twelve hundreds (1200) data from University of Chittagong where about thirty (30000) thousands students are studying. To assess these large amounts of data we have found that SVM is very efficient and exact technique in our proceedings. By imposing the SVM, we have mapped the data to meaning full forty two (42) data which are shown in the figure 3 and 4.In figure 3 we have depicts that the reasons for the age where mainly focused on the pivotal age of greater or less than twenty four.

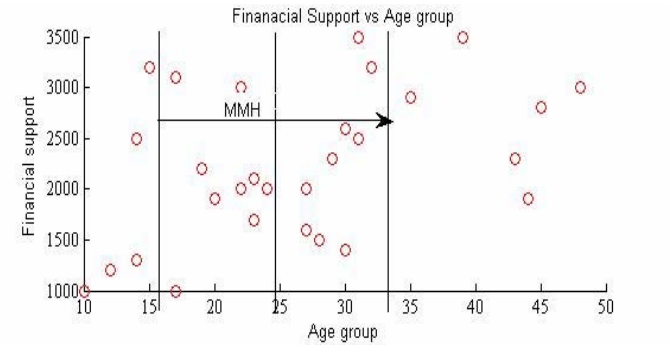


Fig. 3 : Data classification for age group using SVM

From the figure above we can easily measure the Maximum Margin Hyper plane (MMH).At MMH the resultant outcome of age group using SVM is determined. We have also used the MATHLAB to accelerate the accuracy of the implementation. In the same process we have had accomplished our design and implementation for the financial support data using the methodology of SVM.

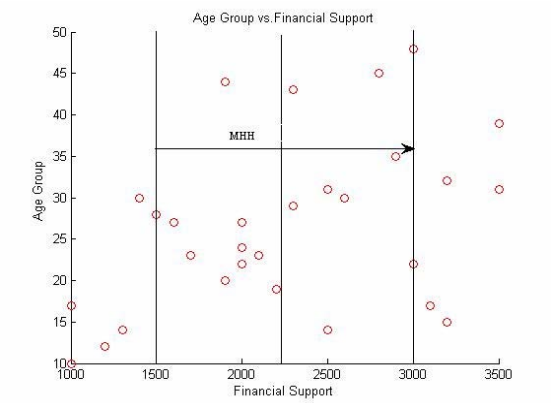


Fig. 4 : SVM to classify the financial support data

Now, we have calculated the chi square values of collected data. The procedures of chi square values are given bellow:

Step1: First insert the observed value in each cell of observable table. Inserted value collected from record.

Domain category	Option1	Option 2	Total
Cotegory1	a	b	a + b
Category	c	d	c + d
Total	a +	b	a + b + c

Step 2: Calculate expected value for every cell of the describing table.

Domain	Option 1	Option 2	Total
Cotegory1	a1=(a+b)*(a+c)/(a+b	b1=(a+b)*(b+d)/(a+b+c	a1 + b1
Category2	c1=(a+c)*(b+d)/(a+b	d1=(c+d)*(b+d)/(a+b+c	c1 + d1
Total	a1 + c1	b1 +d1	a1+

Step 3: calculating chi value for every cell using the following formula:

$\chi^2 = (observed\ value\text{-}expected\ value)^2/expected\ value$

Step 4: calculate total chi –value for domain using the following formula

$$\chi^2 = \sum_{i=1}^{n} (observed\ value\text{-}expected\ value)^2/expected\ value$$

Step 5: calculating degree of freedom using following rule

Degree of freedom,
df = (No.of.rows-1)*(No.of.columns-1)

Step 6: calculate p-value (probability value) using following method in Ms Excel

P-value=CHIDIS (Chi value, df)

VI. Explanation of Chi-Square (χ^2) and P-Value

Step 1: consider the domain is financial support in which category1=Yes category2=No, Option1=Pass and Option2=Fail. The value of every cell collects from database.

Financial Support	Pass	Fail	Total
Yes	18	5	23
No	11	8	19
Total	29	13	42

Step 2: calculating expected value for each cell using describing formula

Financial Support	Pass	Fail	Total
Yes	23*29/42=15.88	23*13/42=7.12	23
No	19*29/42=13.12	19*13/42=5.88	19
Total	29	13	42

Step 3: calculating chi value for every cell using the describing formula:

Financial Support	Pass	Fail
Yes	$\chi^2=(18-15.88)^2/15.88 = 0.283$	$\chi^2=(5-7.12)^2/7.12 = 0.631$
No	$\chi^2=(11-13.12)^2/13.88 = 0.343$	$\chi^2=(8-5.88)^2/5.88 = 0.764$

Step 4: calculate total chi –value for domain

$\chi^2= 0.283+0.631+0.343+0.764=2.021$

Overall chi-value for every domain in Pass and Fail Category:

Table 3 : Individual Chi-value for each category

Domain Name	Possible category	Pass	Fail
Financial Support	Yes	0.283	0.631
	No	0.343	0.764
Age group	Age>24	0.337	0.753
	Age<24	0.120	0.267
Gender	Male	0.050	0.112
	Female	0.082	0.182
Disabilities	Yes	0	1.54
	Female	0.20	0.040

Step 5: calculating degree of freedom using following the rule

Degree of freedom df = (2-1)*(2-1)
= 1

Step 6: calculate p-value (probability value) using in Ms Excels

P-value =CHIDIS (2.021, 1)
= 0.155

VII. Factors Selection

Factors selection is an important process to assess the prediction of dropout of the students. The prediction has relate the variable that determines the rate of success The number of predictor variables is not so large and we don't have to select the subset of variables for further analysis which is the main purpose of applying feature selection to data. However, feature selection could be also used as a pre-processor for predictive data mining to rank predictors according to the strength of their relationship with dependent or outcome variable. During the factors selection process no specific form of relationship, neither linear nor nonlinear, is assumed. The outcome of the factors selection would be a rank list of predictors according to their importance for further analysis of the dependent variable with the other methods for regression and classification. Here the figure below shows the relative outcome of the predictor's value.

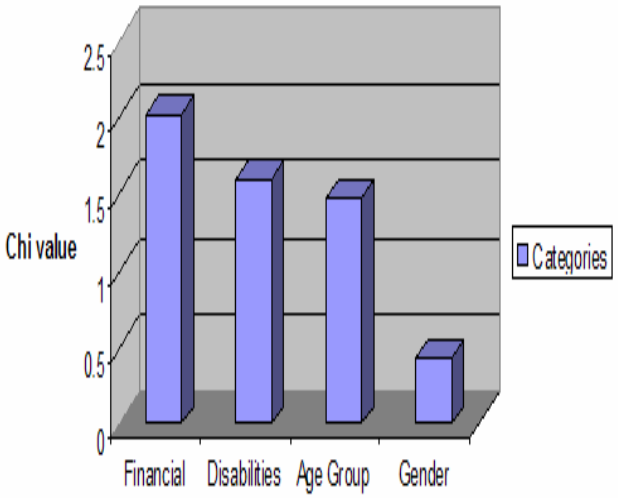


Figure 5 : Importance plot for predictors

Results of factors selection has presented in Figure 3 on the importance plot and in Table 1. The top three predictors for the study outcome are financial support of students, disabilities, age group and gender in which they are study.

Table 4 : Best predictors for dependent variable

Domain Name	Chi-value	P-value
Financial Support	2.021	.155
Disabilities	1.6	.206
Age Group	1.477	.224
Gender	.426	.514

In all three cases, i.e. for all three definitions of the dependent variable, if the top 4 variables are selected, we get the same list of predictors. Therefore we can conclude that the list of important predictors is quite robust to changes in the study outcome definition. We may proceed into the next step using the top 4 variables:

1. Financial Support
2. Age Group
3. Gender
4. Disabilities

From Table 4, *P*-values from the last column only the first three chi-square values are significant at 10% level. Though the results of the feature selection suggested continuing analysis with only the subset of predictors, which includes Financial Support, Age group and Gender, we have included all available predictors in our classification tree analysis. We follow an advice given in Luan & Zhao (2006) who suggested that even though some variables may have little significance to the overall prediction outcome, they can be essential to a specific record.

a) Contribution of Cart

Figure 6 shows the CART classification tree for study outcome. It shows that only three variables were used to construct the tree: (1) financial support (2) age group and (3) gender.

The largest successful group (i.e. students who successfully completed the course) consists of 18 (78.76% of all participants) students (Node 1). The financial support of students in this group is yes. Students in this group enrolled on the age group in either age>24 or age<=24. The largest successful group (i.e. age group) contains 9 (60% of all participants) students (Node 4).

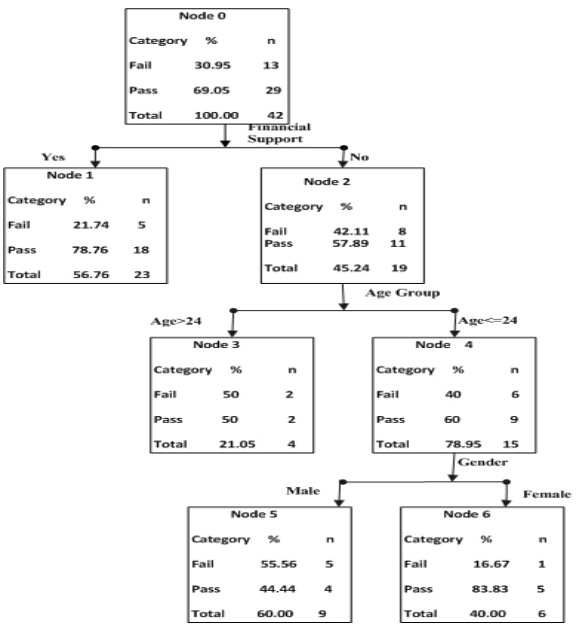


Fig. 6 : CART Classification

VIII. Knowledge Base for Collected Data

A knowledge base in artificial intelligence is a place where information are stored or designed for machine or device by which it will work. In general, a knowledge base is a consolidate stock for information: a library, a database of related information about a particular subject could all be considered to be examples of knowledge bases. The process of building knowledge base is called knowledge engineering. A knowledge base is integrated collection of choosing logic, building a knowledge base, implementing [31] the proof theory, inferring new facts. The main advantage of engineering is that it requires less commitment and thus less work. To help the focus the development of knowledge base and to integrate the designer's thinking the following five step methodology can be used:

1. Decide what to talk about
2. Decide on a vocabulary of predicates, function, and constant.
3. Encode general knowledge about the domain.
4. Encode a description of the specific problem instance.
5. Pose queries to the inference procedure and answers.

In our work we have described a simple method of probabilistic inference that is, the computation from observed evidence of posterior probabilities for query propositions. We have used the joint probability as the knowledge base from which answer to all question may be derived. We have had built the knowledge base by considering two Boolean variables. The table 7 is an example of two valued propositional logic which is the bases of knowledge base representation:

Table 7: Concepts of propositional logic to design a Knowledge Base using the proposition of Boolean events A, B and C.

	B		⌐B	
	C	⌐C	C	⌐C
A	111	110	101	100
⌐A	011	010	001	000

Based on table 7, we have designed the knowledge base (Joint probability distribution) for our research activity. Here we have considered those events which have true (one or 1) Boolean values. Table 8 is an example of knowledge base for events A, B and C:

Table 5 : Fully Joint probability distribution

	B		⌐ B	
	C	⌐ C	C	⌐ C
A	P(A)*P(B)*P(C)	P(A)*P(B)*P(⌐ C)	P(A)*P(⌐ B)*P(C)	P(A)*P(⌐B)*P (C)
⌐ A	P(⌐A)*P(B)*P(C)	P(⌐ A)*P(B)*P(⌐ C)	P(A)*P(⌐ B)*P(C)	P(A)*P(⌐B)*P (C)

By keeping the similarities s with the table 5, we compared our factors as financially good and financially not good, fail and not fail and so on. The designing of knowledge base for the factors which we are considered are given in table 6:

Table 6 : Fully join probability distribution (Knowledge base)

	Financial		⌐ Financial	
	Age<=24	Age>24	Age<=24	Age>24
Fail	23/42*31/42*13/42 =0.125	23/42*11/42*13/42 =0.044	19/42*31/42*13/42 =0.103	11/42*19/42*13/42 =0.037
Pass	23/42*31/42*29/42 =0.279	23/42*11/42*29/42 =0.099	31/42*19/42*29/42 =0.231	11/42*19/42*29/42 =0.082

Where
0.125+0.044+0.103+0.037+0.279+0.099+0.231 +0.082=1

IX. Bayes'Theorem and Conditional Probability

Bayes' theorem and conditional probability are opposite to each other. Given two dependent events A and B. The conditional probability of P (A and B) or P (B/A) will be P (A and B)/P (A). Related to this formula a rule is developed by the English Presbyterian minister Thomas Bayes (1702-61).According to the Bayes rule it is possible to determine the various probabilities of the first event given the outcome of the second event in a sequence of two events.

The conditional probability:

$$P(B/A) = \frac{P(AandB)}{P(A)} \quad (1)$$

The equation (1) will help to find out the probabilities of B after being occurrences of the A. we get the Bayes' theorem for these two events as follows:

$$P(A/B) = \frac{P(A).P(B/A)}{P(B)} \quad (2)$$

If there are more events like A1, A2, and B1, B2.In this case the Bayes theorem to determine the probability of A1 based on B1will be as follows:

$$P(A1/B1) = \frac{P(A1).P(B1/A1)}{P(A1).P(B1/A1) + P(A2).P(B2/A2)}$$

X. Experimental Result

Consideration Classification and Regression Tree (CART):

For Node 4:
IF Financial Support="Yes" AND Age<=24 THEN "PASS"= (18/23)*(9/15) =0.49

Now applying the Bayes theorem on table 5 we have got the following outcomes:

If one student pass based on his financial condition is "Yes" and age<=24 then
P (Pass | Financial condition="Yes" ^age<=24) =

P (Pass ^Financial condition="Yes" ^age<=24)/ P (Financial condition="Yes" ^age<=24)

P (Pass ^Financial condition="Yes" age<=24) =0.279

P (Financial condition="Yes" ^ age<=24) =0.125+0.279 = 0.404

P (Pass | Financial condition="Yes" ^age<=24) =0.125/0.404=0.31

The total resultant of Bayes Theorem and CART of all data considering financial condition and age group we have got the following table 7:

Table 7 : Bayes Rules to predict the academic success

Rule	Study Outcome	Probability	
		CART	Bayes Theorem
IF Financial Support="Yes" AND Age<=24 THEN	PASS	0.47	0.69
IF Financial Support="Yes" AND Age>24	PASS	0.39	0.69
IF Financial Support="NO" AND Age<=24	PASS	0.35	0.68
IF Financial Support="NO" AND Age>24	PASS	0.29	0.68

XI. Conclusion

This study examines the background information from enrolment data that impacts upon the study outcome of Information Systems students at the Department of Computer Science & Engineering. Based on results from feature selection (Figure 5 and Table 4), the CART trees presentation it was found that the most important factors that help separate successful from unsuccessful students are financial support , age group and gender. Demographic data such as gender and age though significantly related to the study outcome, according to the feature selection result.

Based on results from table 5 and 6 by implementing the knowledge of propositional knowledge base and Bayes theorem based on knowledge base to predict the academic success. Better result is found from using Bayes Network.

This study is limited in three main ways that future research can perhaps address. Firstly, this research is based on background information only. Secondly, we used a dichotomous variable for the study outcome with only two categories: pass and fail. Thirdly, from a methodological point of view an alternative to a classification tree should be considered. The prime candidates to be used with this data set are logistic regression and neural networks.

References Références Referencias

1. Al-Radaideh, Q. A., Al-Shawakfa, E. M., & Al-Najjar, M. I. (2006). Mining student data using decision trees. In the Proceedings of the 2006 International Arab Conference on Information Technology (ACIT'2006*)*.
2. Md.Sarwar Kamal, Linkon Chowdhury, Sonia Farhana Nimmy, *"*New Dropout Prediction for Intelligent System*", International Journal of Computer Applications (0975 – 8887) Volume:42– No.16, March 2012.*
3. Bean, J. P., & Metzner, B. S. (1985). A conceptual model of non traditional undergraduate student attrition. *Review of Educational Research, 55, 485-540.*
4. Boero, G., Laureti, T., & Naylor, R. (2005). An econometric analysis of student withdrawal and progress- sion in post-reform Italian universities. Centro Ricerche Economiche Nord Sud - CRENoS Working Paper 2005/04.
5. Cortez, P., & Silva, A. (2008). Using data mining to predict secondary school student performance. In the Proceedings of 5th Annual Future Business Technology Conference, Porto, Portugal, 5-12. Dekker, G. W., Pechenizkiy, M., & Vleeshouwers, J. M. (2009). Predicting student drop out: A case study.
6. Proceedings of the 2nd International Conference on Educational Data Mining (EDM'09). July 1-3, Cordoba, Spain, 41-50.
7. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data* mining, infer- ence and prediction (2nd ed.). New York: Springer.
8. Han, J., & Kamber, M. (2006). Data mining: Concepts and techniques (2nd ed.). Amsterdam: Elsevier. Herrera, O. L. (2006). Investigation of the role of pre- and post admission variables in undergraduate insti-Mutational persistence, using a Markov student flow model. PhD Dissertation, North Carolina State Uni-versity, USA.
9. Horstmanshof, L., & Zimitat, C. (2007). Future time orientation predicts academic engagement among first-year university students. British Journal of Educational Psychology, *77* (3): 703-718.
10. Ishitani, T. T. (2003). A longitudinal approach to assessing attrition behavior among first-generation stu- dents: Time-varying effects of precollege characteristics. *Research in Higher Education, 44(4), 433-449.*
11. Ishitani, T. T. (2006). Studying attrition and degree completion behavior among first-generation college students in the United States. Journal of Higher Education, 77(5), 861-885.
12. Jun, J. (2005). Understanding dropout of adult learners in e-learning. PhD Dissertation, the University of Georgia, USA.
13. Kember, D. (1995). *Open learning courses for adults: A model of student progress*. Englewood Cliffs, NJ: Education Technology.
14. Luan, J., & Zhao, C-M. (2006). Practicing data mining for enrolment management and beyond. New Di- rections for Institutional Research, 31(1), 117-122.
15. Murtaugh, P., Burns, L., & Schuster, J. (1999). Predicting the retention of university students. Research in Higher Education, 40(3), 355- 371.
16. Nandeshwar, A., & Chaudhari, S. (2009). Enrollment prediction models using data mining. Retrieved Janu- ary 10, 2010,
17. Nisbet, R., Elder, J., & Miner, G. (2009). Handbook of statistical analysis and data mining applications. Amsterdam: Elsevier.
18. Noble, K., Flynn, N. T., Lee, J. D., & Hilton, D. (2007). Predicting successful college experiences: Evi- dence from a first year retention program. Journal of College Student Retention: Research, Theory & Practice, 9(1), 39-60.
19. Pascarella, E. T., Duby, P. B., & Iverson, B. K. (1983). A test and reconceptualization of a theoretical model of college withdrawal in a commuter institution setting. Sociology of Education, 56, 88-100.
20. Pratt, P. A., & Skaggs, C. T. (1989). First-generation college students: Are they at greater risk for attrition

than their peers? *Research in Rural Education, 6* (1), 31-34.
21. Reason, R. D. (2003). Student variables that predict retention: Recent research and new developments. *NASPA Journal, 40(4),* 172-191.
22. Rokach, L., & Maimon, O. (2008). Data mining with decision trees – Theory and applications. New Jersey: World Scientific Publishing.
23. Yu, C. H., DiGangi, S., Jannasch-Pennell, A., Lo, W., & Kaprolet, C. (2007). A data-mining approach to differentiate predictors of retention. In the Proceedings of the Educause Southwest Conference, Austin, Texas, USA.
24. Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. Expert Systems with Applications, 33, 135-146.
25. Simpson, O. (2006). Predicting student success in open and distance learning. *Open Learning, 21*(2), 125-138.
26. Siraj, F., & Abdoulha, M. A. (2009). Uncovering hidden information within university's student enrolment data using data mining. MASAUM Journal of Computing, 1(2), 337-342.
27. Strayhorn, T. L. (2009). An examination of the impact of first-year seminars on correlates of college stu- dent retention. Journal of the First- Year Experience & Students in Transition, 21(1), 9-27.
28. Tharp, J. (1998). Predicting persistence of urban commuter campus students utilizing student background characteristics from enrollment data. Community College Journal of Research and Practice, 22, 279-294.
29. Artificial Intelligence A modern Approach by Stuart Russell and Peter Norvig, ISBN 81-297-0041-7
30. Baker, R. S. J. D., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. Journal of Educational Data Mining, 1, 3-17.

Global Journals Inc. (US) Guidelines Handbook 2012

www.GlobalJournals.org

FELLOW OF ASSOCIATION OF RESEARCH SOCIETY IN COMPUTING (FARSC)

- 'FARSC' title will be awarded to the person after approval of Editor-in-Chief and Editorial Board. The title 'FARSC" can be added to name in the following manner. eg. **Dr. John E. Hall, Ph.D., FARSC or William Walldroff Ph. D., M.S., FARSC**
- Being FARSC is a respectful honor. It authenticates your research activities. After becoming FARSC, you can use 'FARSC' title as you use your degree in suffix of your name. This will definitely will enhance and add up your name. You can use it on your Career Counseling Materials/CV/Resume/Visiting Card/Name Plate etc.
- 60% Discount will be provided to FARSC members for publishing research papers in Global Journals Inc., if our Editorial Board and Peer Reviewers accept the paper. For the life time, if you are author/co-author of any paper bill sent to you will automatically be discounted one by 60%
- FARSC will be given a renowned, secure, free professional email address with 100 GB of space eg.johnhall@globaljournals.org. You will be facilitated with Webmail, Spam Assassin, Email Forwarders, Auto-Responders, Email Delivery Route tracing, etc.
- FARSC member is eligible to become paid peer reviewer at Global Journals Inc. to earn up to 15% of realized author charges taken from author of respective paper. After reviewing 5 or more papers you can request to transfer the amount to your bank account or to your PayPal account.
- Eg. If we had taken 420 USD from author, we can send 63 USD to your account.
- FARSC member can apply for free approval, grading and certification of some of their Educational and Institutional Degrees from Global Journals Inc. (US) and Open Association of Research, Society U.S.A.
- After you are FARSC. You can send us scanned copy of all of your documents. We will verify, grade and certify them within a month. It will be based on your academic records, quality of research papers published by you, and 50 more criteria. This is beneficial for your job interviews as recruiting organization need not just rely on you for authenticity and your unknown qualities, you would have authentic ranks of all of your documents. Our scale is unique worldwide.
- FARSC member can proceed to get benefits of free research podcasting in Global Research Radio with their research documents, slides and online movies.
- After your publication anywhere in the world, you can upload you research paper with your recorded voice or you can use our professional RJs to record your paper their voice. We can also stream your conference videos and display your slides online.
- FARSC will be eligible for free application of Standardization of their Researches by Open Scientific Standards. Standardization is next step and level after publishing in a journal. A team of research and professional will work with you to take your research to its next level, which is worldwide open standardization.

- FARSC is eligible to earn from their researches: While publishing his paper with Global Journals Inc. (US), FARSC can decide whether he/she would like to publish his/her research in closed manner. When readers will buy that individual research paper for reading, 80% of its earning by Global Journals Inc. (US) will be transferred to FARSC member's bank account after certain threshold balance. There is no time limit for collection. FARSC member can decide its price and we can help in decision.

MEMBER OF ASSOCIATION OF RESEARCH SOCIETY IN COMPUTING (MARSC)

- 'MARSC' title will be awarded to the person after approval of Editor-in-Chief and Editorial Board. The title 'MARSC" can be added to name in the following manner. eg. Dr. John E. Hall, Ph.D., MARSC or William Walldroff Ph. D., M.S., MARSC
- Being MARSC is a respectful honor. It authenticates your research activities. After becoming MARSC, you can use 'MARSC' title as you use your degree in suffix of your name. This will definitely will enhance and add up your name. You can use it on your Career Counseling Materials/CV/Resume/Visiting Card/Name Plate etc.
- 40% Discount will be provided to MARSC members for publishing research papers in Global Journals Inc., if our Editorial Board and Peer Reviewers accept the paper. For the life time, if you are author/co-author of any paper bill sent to you will automatically be discounted one by 60%
- MARSC will be given a renowned, secure, free professional email address with 30 GB of space eg.johnhall@globaljournals.org. You will be facilitated with Webmail, Spam Assassin, Email Forwarders, Auto-Responders, Email Delivery Route tracing, etc.
- MARSC member is eligible to become paid peer reviewer at Global Journals Inc. to earn up to 10% of realized author charges taken from author of respective paper. After reviewing 5 or more papers you can request to transfer the amount to your bank account or to your PayPal account.
- MARSC member can apply for free approval, grading and certification of some of their Educational and Institutional Degrees from Global Journals Inc. (US) and Open Association of Research, Society U.S.A.
- MARSC is eligible to earn from their researches: While publishing his paper with Global Journals Inc. (US), MARSC can decide whether he/she would like to publish his/her research in closed manner. When readers will buy that individual research paper for reading, 40% of its earning by Global Journals Inc. (US) will be transferred to MARSC member's bank account after certain threshold balance. There is no time limit for collection. MARSC member can decide its price and we can help in decision.

AUXILIARY MEMBERSHIPS

ANNUAL MEMBER

- Annual Member will be authorized to receive e-Journal GJCST for one year (subscription for one year).
- The member will be allotted free 1 GB Web-space along with subDomain to contribute and participate in our activities.
- A professional email address will be allotted free 500 MB email space.

PAPER PUBLICATION

- The members can publish paper once. The paper will be sent to two-peer reviewer. The paper will be published after the acceptance of peer reviewers and Editorial Board.

Process of submission of Research Paper

The Area or field of specialization may or may not be of any category as mentioned in 'Scope of Journal' menu of the GlobalJournals.org website. There are 37 Research Journal categorized with Six parental Journals GJCST, GJMR, GJRE, GJMBR, GJSFR, GJHSS. For Authors should prefer the mentioned categories. There are three widely used systems UDC, DDC and LCC. The details are available as 'Knowledge Abstract' at Home page. The major advantage of this coding is that, the research work will be exposed to and shared with all over the world as we are being abstracted and indexed worldwide.

The paper should be in proper format. The format can be downloaded from first page of 'Author Guideline' Menu. The Author is expected to follow the general rules as mentioned in this menu. The paper should be written in MS-Word Format (*.DOC,*.DOCX).

The Author can submit the paper either online or offline. The authors should prefer online submission.Online Submission: There are three ways to submit your paper:

(A) (I) First, register yourself using top right corner of Home page then Login. If you are already registered, then login using your username and password.

(II) Choose corresponding Journal.

(III) Click 'Submit Manuscript'. Fill required information and Upload the paper.

(B) If you are using Internet Explorer, then Direct Submission through Homepage is also available.

(C) If these two are not convenient, and then email the paper directly to dean@globaljournals.org.

Offline Submission: Author can send the typed form of paper by Post. However, online submission should be preferred.

Preferred Author Guidelines

MANUSCRIPT STYLE INSTRUCTION (Must be strictly followed)

Page Size: 8.27" X 11'"

- Left Margin: 0.65
- Right Margin: 0.65
- Top Margin: 0.75
- Bottom Margin: 0.75
- Font type of all text should be Swis 721 Lt BT.
- Paper Title should be of Font Size 24 with one Column section.
- Author Name in Font Size of 11 with one column as of Title.
- Abstract Font size of 9 Bold, "Abstract" word in Italic Bold.
- Main Text: Font size 10 with justified two columns section
- Two Column with Equal Column with of 3.38 and Gaping of .2
- First Character must be three lines Drop capped.
- Paragraph before Spacing of 1 pt and After of 0 pt.
- Line Spacing of 1 pt
- Large Images must be in One Column
- Numbering of First Main Headings (Heading 1) must be in Roman Letters, Capital Letter, and Font Size of 10.
- Numbering of Second Main Headings (Heading 2) must be in Alphabets, Italic, and Font Size of 10.

You can use your own standard format also.
Author Guidelines:

1. General,

2. Ethical Guidelines,

3. Submission of Manuscripts,

4. Manuscript's Category,

5. Structure and Format of Manuscript,

6. After Acceptance.

1. GENERAL

Before submitting your research paper, one is advised to go through the details as mentioned in following heads. It will be beneficial, while peer reviewer justify your paper for publication.

Scope

The Global Journals Inc. (US) welcome the submission of original paper, review paper, survey article relevant to the all the streams of Philosophy and knowledge. The Global Journals Inc. (US) is parental platform for Global Journal of Computer Science and Technology, Researches in Engineering, Medical Research, Science Frontier Research, Human Social Science, Management, and Business organization. The choice of specific field can be done otherwise as following in Abstracting and Indexing Page on this Website. As the all Global

Journals Inc. (US) are being abstracted and indexed (in process) by most of the reputed organizations. Topics of only narrow interest will not be accepted unless they have wider potential or consequences.

2. ETHICAL GUIDELINES

Authors should follow the ethical guidelines as mentioned below for publication of research paper and research activities.

Papers are accepted on strict understanding that the material in whole or in part has not been, nor is being, considered for publication elsewhere. If the paper once accepted by Global Journals Inc. (US) and Editorial Board, will become the copyright of the Global Journals Inc. (US).

Authorship: The authors and coauthors should have active contribution to conception design, analysis and interpretation of findings. They should critically review the contents and drafting of the paper. All should approve the final version of the paper before submission

The Global Journals Inc. (US) follows the definition of authorship set up by the Global Academy of Research and Development. According to the Global Academy of R&D authorship, criteria must be based on:

1) Substantial contributions to conception and acquisition of data, analysis and interpretation of the findings.

2) Drafting the paper and revising it critically regarding important academic content.

3) Final approval of the version of the paper to be published.

All authors should have been credited according to their appropriate contribution in research activity and preparing paper. Contributors who do not match the criteria as authors may be mentioned under Acknowledgement.

Acknowledgements: Contributors to the research other than authors credited should be mentioned under acknowledgement. The specifications of the source of funding for the research if appropriate can be included. Suppliers of resources may be mentioned along with address.

Appeal of Decision: The Editorial Board's decision on publication of the paper is final and cannot be appealed elsewhere.

Permissions: It is the author's responsibility to have prior permission if all or parts of earlier published illustrations are used in this paper.

Please mention proper reference and appropriate acknowledgements wherever expected.

If all or parts of previously published illustrations are used, permission must be taken from the copyright holder concerned. It is the author's responsibility to take these in writing.

Approval for reproduction/modification of any information (including figures and tables) published elsewhere must be obtained by the authors/copyright holders before submission of the manuscript. Contributors (Authors) are responsible for any copyright fee involved.

3. SUBMISSION OF MANUSCRIPTS

Manuscripts should be uploaded via this online submission page. The online submission is most efficient method for submission of papers, as it enables rapid distribution of manuscripts and consequently speeds up the review procedure. It also enables authors to know the status of their own manuscripts by emailing us. Complete instructions for submitting a paper is available below.

Manuscript submission is a systematic procedure and little preparation is required beyond having all parts of your manuscript in a given format and a computer with an Internet connection and a Web browser. Full help and instructions are provided on-screen. As an author, you will be prompted for login and manuscript details as Field of Paper and then to upload your manuscript file(s) according to the instructions.

To avoid postal delays, all transaction is preferred by e-mail. A finished manuscript submission is confirmed by e-mail immediately and your paper enters the editorial process with no postal delays. When a conclusion is made about the publication of your paper by our Editorial Board, revisions can be submitted online with the same procedure, with an occasion to view and respond to all comments.

Complete support for both authors and co-author is provided.

4. MANUSCRIPT'S CATEGORY

Based on potential and nature, the manuscript can be categorized under the following heads:

Original research paper: Such papers are reports of high-level significant original research work.

Review papers: These are concise, significant but helpful and decisive topics for young researchers.

Research articles: These are handled with small investigation and applications

Research letters: The letters are small and concise comments on previously published matters.

5.STRUCTURE AND FORMAT OF MANUSCRIPT

The recommended size of original research paper is less than seven thousand words, review papers fewer than seven thousands words also.Preparation of research paper or how to write research paper, are major hurdle, while writing manuscript. The research articles and research letters should be fewer than three thousand words, the structure original research paper; sometime review paper should be as follows:

Papers: These are reports of significant research (typically less than 7000 words equivalent, including tables, figures, references), and comprise:

(a)Title should be relevant and commensurate with the theme of the paper.

(b) A brief Summary, "Abstract" (less than 150 words) containing the major results and conclusions.

(c) Up to ten keywords, that precisely identifies the paper's subject, purpose, and focus.

(d) An Introduction, giving necessary background excluding subheadings; objectives must be clearly declared.

(e) Resources and techniques with sufficient complete experimental details (wherever possible by reference) to permit repetition; sources of information must be given and numerical methods must be specified by reference, unless non-standard.

(f) Results should be presented concisely, by well-designed tables and/or figures; the same data may not be used in both; suitable statistical data should be given. All data must be obtained with attention to numerical detail in the planning stage. As reproduced design has been recognized to be important to experiments for a considerable time, the Editor has decided that any paper that appears not to have adequate numerical treatments of the data will be returned un-refereed;

(g) Discussion should cover the implications and consequences, not just recapitulating the results; conclusions should be summarizing.

(h) Brief Acknowledgements.

(i) References in the proper form.

Authors should very cautiously consider the preparation of papers to ensure that they communicate efficiently. Papers are much more likely to be accepted, if they are cautiously designed and laid out, contain few or no errors, are summarizing, and be conventional to the approach and instructions. They will in addition, be published with much less delays than those that require much technical and editorial correction.

The Editorial Board reserves the right to make literary corrections and to make suggestions to improve briefness.

It is vital, that authors take care in submitting a manuscript that is written in simple language and adheres to published guidelines.

Format

Language: The language of publication is UK English. Authors, for whom English is a second language, must have their manuscript efficiently edited by an English-speaking person before submission to make sure that, the English is of high excellence. It is preferable, that manuscripts should be professionally edited.

Standard Usage, Abbreviations, and Units: Spelling and hyphenation should be conventional to The Concise Oxford English Dictionary. Statistics and measurements should at all times be given in figures, e.g. 16 min, except for when the number begins a sentence. When the number does not refer to a unit of measurement it should be spelt in full unless, it is 160 or greater.

Abbreviations supposed to be used carefully. The abbreviated name or expression is supposed to be cited in full at first usage, followed by the conventional abbreviation in parentheses.

Metric SI units are supposed to generally be used excluding where they conflict with current practice or are confusing. For illustration, 1.4 l rather than 1.4 × 10-3 m3, or 4 mm somewhat than 4 × 10-3 m. Chemical formula and solutions must identify the form used, e.g. anhydrous or hydrated, and the concentration must be in clearly defined units. Common species names should be followed by underlines at the first mention. For following use the generic name should be constricted to a single letter, if it is clear.

Structure

All manuscripts submitted to Global Journals Inc. (US), ought to include:

Title: The title page must carry an instructive title that reflects the content, a running title (less than 45 characters together with spaces), names of the authors and co-authors, and the place(s) wherever the work was carried out. The full postal address in addition with the e-mail address of related author must be given. Up to eleven keywords or very brief phrases have to be given to help data retrieval, mining and indexing.

Abstract, used in Original Papers and Reviews:

Optimizing Abstract for Search Engines

Many researchers searching for information online will use search engines such as Google, Yahoo or similar. By optimizing your paper for search engines, you will amplify the chance of someone finding it. This in turn will make it more likely to be viewed and/or cited in a further work. Global Journals Inc. (US) have compiled these guidelines to facilitate you to maximize the web-friendliness of the most public part of your paper.

Key Words

A major linchpin in research work for the writing research paper is the keyword search, which one will employ to find both library and Internet resources.

One must be persistent and creative in using keywords. An effective keyword search requires a strategy and planning a list of possible keywords and phrases to try.

Search engines for most searches, use Boolean searching, which is somewhat different from Internet searches. The Boolean search uses "operators," words (and, or, not, and near) that enable you to expand or narrow your affords. Tips for research paper while preparing research paper are very helpful guideline of research paper.

Choice of key words is first tool of tips to write research paper. Research paper writing is an art.A few tips for deciding as strategically as possible about keyword search:

- One should start brainstorming lists of possible keywords before even begin searching. Think about the most important concepts related to research work. Ask, "What words would a source have to include to be truly valuable in research paper?" Then consider synonyms for the important words.
- It may take the discovery of only one relevant paper to let steer in the right keyword direction because in most databases, the keywords under which a research paper is abstracted are listed with the paper.
- One should avoid outdated words.

Keywords are the key that opens a door to research work sources. Keyword searching is an art in which researcher's skills are bound to improve with experience and time.

Numerical Methods: Numerical methods used should be clear and, where appropriate, supported by references.

Acknowledgements: Please make these as concise as possible.

References

References follow the Harvard scheme of referencing. References in the text should cite the authors' names followed by the time of their publication, unless there are three or more authors when simply the first author's name is quoted followed by et al. unpublished work has to only be cited where necessary, and only in the text. Copies of references in press in other journals have to be supplied with submitted typescripts. It is necessary that all citations and references be carefully checked before submission, as mistakes or omissions will cause delays.

References to information on the World Wide Web can be given, but only if the information is available without charge to readers on an official site. Wikipedia and Similar websites are not allowed where anyone can change the information. Authors will be asked to make available electronic copies of the cited information for inclusion on the Global Journals Inc. (US) homepage at the judgment of the Editorial Board.

The Editorial Board and Global Journals Inc. (US) recommend that, citation of online-published papers and other material should be done via a DOI (digital object identifier). If an author cites anything, which does not have a DOI, they run the risk of the cited material not being noticeable.

The Editorial Board and Global Journals Inc. (US) recommend the use of a tool such as Reference Manager for reference management and formatting.

Tables, Figures and Figure Legends

Tables: Tables should be few in number, cautiously designed, uncrowned, and include only essential data. Each must have an Arabic number, e.g. Table 4, a self-explanatory caption and be on a separate sheet. Vertical lines should not be used.

Figures: Figures are supposed to be submitted as separate files. Always take in a citation in the text for each figure using Arabic numbers, e.g. Fig. 4. Artwork must be submitted online in electronic form by e-mailing them.

Preparation of Electronic Figures for Publication

Even though low quality images are sufficient for review purposes, print publication requires high quality images to prevent the final product being blurred or fuzzy. Submit (or e-mail) EPS (line art) or TIFF (halftone/photographs) files only. MS PowerPoint and Word Graphics are unsuitable for printed pictures. Do not use pixel-oriented software. Scans (TIFF only) should have a resolution of at least 350 dpi (halftone) or 700 to 1100 dpi (line drawings) in relation to the imitation size. Please give the data for figures in black and white or submit a Color Work Agreement Form. EPS files must be saved with fonts embedded (and with a TIFF preview, if possible).

For scanned images, the scanning resolution (at final image size) ought to be as follows to ensure good reproduction: line art: >650 dpi; halftones (including gel photographs) : >350 dpi; figures containing both halftone and line images: >650 dpi.

Color Charges: It is the rule of the Global Journals Inc. (US) for authors to pay the full cost for the reproduction of their color artwork. Hence, please note that, if there is color artwork in your manuscript when it is accepted for publication, we would require you to complete and return a color work agreement form before your paper can be published.

Figure Legends: Self-explanatory legends of all figures should be incorporated separately under the heading 'Legends to Figures'. In the full-text online edition of the journal, figure legends may possibly be truncated in abbreviated links to the full screen version. Therefore, the first 100 characters of any legend should notify the reader, about the key aspects of the figure.

6. AFTER ACCEPTANCE

Upon approval of a paper for publication, the manuscript will be forwarded to the dean, who is responsible for the publication of the Global Journals Inc. (US).

6.1 Proof Corrections

The corresponding author will receive an e-mail alert containing a link to a website or will be attached. A working e-mail address must therefore be provided for the related author.

Acrobat Reader will be required in order to read this file. This software can be downloaded

(Free of charge) from the following website:

www.adobe.com/products/acrobat/readstep2.html. This will facilitate the file to be opened, read on screen, and printed out in order for any corrections to be added. Further instructions will be sent with the proof.

Proofs must be returned to the dean at dean@globaljournals.org within three days of receipt.

As changes to proofs are costly, we inquire that you only correct typesetting errors. All illustrations are retained by the publisher. Please note that the authors are responsible for all statements made in their work, including changes made by the copy editor.

6.2 Early View of Global Journals Inc. (US) (Publication Prior to Print)

The Global Journals Inc. (US) are enclosed by our publishing's Early View service. Early View articles are complete full-text articles sent in advance of their publication. Early View articles are absolute and final. They have been completely reviewed, revised and edited for publication, and the authors' final corrections have been incorporated. Because they are in final form, no changes can be made after sending them. The nature of Early View articles means that they do not yet have volume, issue or page numbers, so Early View articles cannot be cited in the conventional way.

6.3 Author Services

Online production tracking is available for your article through Author Services. Author Services enables authors to track their article - once it has been accepted - through the production process to publication online and in print. Authors can check the status of their articles online and choose to receive automated e-mails at key stages of production. The authors will receive an e-mail with a unique link that enables them to register and have their article automatically added to the system. Please ensure that a complete e-mail address is provided when submitting the manuscript.

6.4 Author Material Archive Policy

Please note that if not specifically requested, publisher will dispose off hardcopy & electronic information submitted, after the two months of publication. If you require the return of any information submitted, please inform the Editorial Board or dean as soon as possible.

6.5 Offprint and Extra Copies

A PDF offprint of the online-published article will be provided free of charge to the related author, and may be distributed according to the Publisher's terms and conditions. Additional paper offprint may be ordered by emailing us at: editor@globaljournals.org .

the search? Will I be able to find all information in this field area? If the answer of these types of questions will be "Yes" then you can choose that topic. In most of the cases, you may have to conduct the surveys and have to visit several places because this field is related to Computer Science and Information Technology. Also, you may have to do a lot of work to find all rise and falls regarding the various data of that subject. Sometimes, detailed information plays a vital role, instead of short information.

2. Evaluators are human: First thing to remember that evaluators are also human being. They are not only meant for rejecting a paper. They are here to evaluate your paper. So, present your Best.

3. Think Like Evaluators: If you are in a confusion or getting demotivated that your paper will be accepted by evaluators or not, then think and try to evaluate your paper like an Evaluator. Try to understand that what an evaluator wants in your research paper and automatically you will have your answer.

4. Make blueprints of paper: The outline is the plan or framework that will help you to arrange your thoughts. It will make your paper logical. But remember that all points of your outline must be related to the topic you have chosen.

5. Ask your Guides: If you are having any difficulty in your research, then do not hesitate to share your difficulty to your guide (if you have any). They will surely help you out and resolve your doubts. If you can't clarify what exactly you require for your work then ask the supervisor to help you with the alternative. He might also provide you the list of essential readings.

6. Use of computer is recommended: As you are doing research in the field of Computer Science, then this point is quite obvious.

7. Use right software: Always use good quality software packages. If you are not capable to judge good software then you can lose quality of your paper unknowingly. There are various software programs available to help you, which you can get through Internet.

8. Use the Internet for help: An excellent start for your paper can be by using the Google. It is an excellent search engine, where you can have your doubts resolved. You may also read some answers for the frequent question how to write my research paper or find model research paper. From the internet library you can download books. If you have all required books make important reading selecting and analyzing the specified information. Then put together research paper sketch out.

9. Use and get big pictures: Always use encyclopedias, Wikipedia to get pictures so that you can go into the depth.

10. Bookmarks are useful: When you read any book or magazine, you generally use bookmarks, right! It is a good habit, which helps to not to lose your continuity. You should always use bookmarks while searching on Internet also, which will make your search easier.

11. Revise what you wrote: When you write anything, always read it, summarize it and then finalize it.

12. Make all efforts: Make all efforts to mention what you are going to write in your paper. That means always have a good start. Try to mention everything in introduction, that what is the need of a particular research paper. Polish your work by good skill of writing and always give an evaluator, what he wants.

13. Have backups: When you are going to do any important thing like making research paper, you should always have backup copies of it either in your computer or in paper. This will help you to not to lose any of your important.

14. Produce good diagrams of your own: Always try to include good charts or diagrams in your paper to improve quality. Using several and unnecessary diagrams will degrade the quality of your paper by creating "hotchpotch." So always, try to make and include those diagrams, which are made by your own to improve readability and understandability of your paper.

15. Use of direct quotes: When you do research relevant to literature, history or current affairs then use of quotes become essential but if study is relevant to science then use of quotes is not preferable.

16. Use proper verb tense: Use proper verb tenses in your paper. Use past tense, to present those events that happened. Use present tense to indicate events that are going on. Use future tense to indicate future happening events. Use of improper and wrong tenses will confuse the evaluator. Avoid the sentences that are incomplete.

17. Never use online paper: If you are getting any paper on Internet, then never use it as your research paper because it might be possible that evaluator has already seen it or maybe it is outdated version.

18. **Pick a good study spot:** To do your research studies always try to pick a spot, which is quiet. Every spot is not for studies. Spot that suits you choose it and proceed further.

19. Know what you know: Always try to know, what you know by making objectives. Else, you will be confused and cannot achieve your target.

20. Use good quality grammar: Always use a good quality grammar and use words that will throw positive impact on evaluator. Use of good quality grammar does not mean to use tough words, that for each word the evaluator has to go through dictionary. Do not start sentence with a conjunction. Do not fragment sentences. Eliminate one-word sentences. Ignore passive voice. Do not ever use a big word when a diminutive one would suffice. Verbs have to be in agreement with their subjects. Prepositions are not expressions to finish sentences with. It is incorrect to ever divide an infinitive. Avoid clichés like the disease. Also, always shun irritating alliteration. Use language that is simple and straight forward. put together a neat summary.

21. Arrangement of information: Each section of the main body should start with an opening sentence and there should be a changeover at the end of the section. Give only valid and powerful arguments to your topic. You may also maintain your arguments with records.

22. Never start in last minute: Always start at right time and give enough time to research work. Leaving everything to the last minute will degrade your paper and spoil your work.

23. Multitasking in research is not good: Doing several things at the same time proves bad habit in case of research activity. Research is an area, where everything has a particular time slot. Divide your research work in parts and do particular part in particular time slot.

24. Never copy others' work: Never copy others' work and give it your name because if evaluator has seen it anywhere you will be in trouble.

25. Take proper rest and food: No matter how many hours you spend for your research activity, if you are not taking care of your health then all your efforts will be in vain. For a quality research, study is must, and this can be done by taking proper rest and food.

26. Go for seminars: Attend seminars if the topic is relevant to your research area. Utilize all your resources.

27. Refresh your mind after intervals: Try to give rest to your mind by listening to soft music or by sleeping in intervals. This will also improve your memory.

28. Make colleagues: Always try to make colleagues. No matter how sharper or intelligent you are, if you make colleagues you can have several ideas, which will be helpful for your research.

29. **Think technically:** Always think technically. If anything happens, then search its reasons, its benefits, and demerits.

30. Think and then print: When you will go to print your paper, notice that tables are not be split, headings are not detached from their descriptions, and page sequence is maintained.

31. Adding unnecessary information: Do not add unnecessary information, like, I have used MS Excel to draw graph. Do not add irrelevant and inappropriate material. These all will create superfluous. Foreign terminology and phrases are not apropos. One should NEVER take a broad view. Analogy in script is like feathers on a snake. Not at all use a large word when a very small one would be

sufficient. Use words properly, regardless of how others use them. Remove quotations. Puns are for kids, not grunt readers. Amplification is a billion times of inferior quality than sarcasm.

32. Never oversimplify everything: To add material in your research paper, never go for oversimplification. This will definitely irritate the evaluator. Be more or less specific. Also too, by no means, ever use rhythmic redundancies. Contractions aren't essential and shouldn't be there used. Comparisons are as terrible as clichés. Give up ampersands and abbreviations, and so on. Remove commas, that are, not necessary. Parenthetical words however should be together with this in commas. Understatement is all the time the complete best way to put onward earth-shaking thoughts. Give a detailed literary review.

33. Report concluded results: Use concluded results. From raw data, filter the results and then conclude your studies based on measurements and observations taken. Significant figures and appropriate number of decimal places should be used. Parenthetical remarks are prohibitive. Proofread carefully at final stage. In the end give outline to your arguments. Spot out perspectives of further study of this subject. Justify your conclusion by at the bottom of them with sufficient justifications and examples.

34. After conclusion: Once you have concluded your research, the next most important step is to present your findings. Presentation is extremely important as it is the definite medium though which your research is going to be in print to the rest of the crowd. Care should be taken to categorize your thoughts well and present them in a logical and neat manner. A good quality research paper format is essential because it serves to highlight your research paper and bring to light all necessary aspects in your research.

Informal Guidelines of Research Paper Writing

Key points to remember:

- Submit all work in its final form.
- Write your paper in the form, which is presented in the guidelines using the template.
- Please note the criterion for grading the final paper by peer-reviewers.

Final Points:

A purpose of organizing a research paper is to let people to interpret your effort selectively. The journal requires the following sections, submitted in the order listed, each section to start on a new page.

The introduction will be compiled from reference matter and will reflect the design processes or outline of basis that direct you to make study. As you will carry out the process of study, the method and process section will be constructed as like that. The result segment will show related statistics in nearly sequential order and will direct the reviewers next to the similar intellectual paths throughout the data that you took to carry out your study. The discussion section will provide understanding of the data and projections as to the implication of the results. The use of good quality references all through the paper will give the effort trustworthiness by representing an alertness of prior workings.

Writing a research paper is not an easy job no matter how trouble-free the actual research or concept. Practice, excellent preparation, and controlled record keeping are the only means to make straightforward the progression.

General style:

Specific editorial column necessities for compliance of a manuscript will always take over from directions in these general guidelines.

To make a paper clear

· Adhere to recommended page limits

Mistakes to evade

- Insertion a title at the foot of a page with the subsequent text on the next page

- Separating a table/chart or figure - impound each figure/table to a single page
- Submitting a manuscript with pages out of sequence

In every sections of your document

· Use standard writing style including articles ("a", "the," etc.)

· Keep on paying attention on the research topic of the paper

· Use paragraphs to split each significant point (excluding for the abstract)

· Align the primary line of each section

· Present your points in sound order

· Use present tense to report well accepted

· Use past tense to describe specific results

· Shun familiar wording, don't address the reviewer directly, and don't use slang, slang language, or superlatives

· Shun use of extra pictures - include only those figures essential to presenting results

Title Page:

Choose a revealing title. It should be short. It should not have non-standard acronyms or abbreviations. It should not exceed two printed lines. It should include the name(s) and address (es) of all authors.

Abstract:

The summary should be two hundred words or less. It should briefly and clearly explain the key findings reported in the manuscript--must have precise statistics. It should not have abnormal acronyms or abbreviations. It should be logical in itself. Shun citing references at this point.

An abstract is a brief distinct paragraph summary of finished work or work in development. In a minute or less a reviewer can be taught the foundation behind the study, common approach to the problem, relevant results, and significant conclusions or new questions.

Write your summary when your paper is completed because how can you write the summary of anything which is not yet written? Wealth of terminology is very essential in abstract. Yet, use comprehensive sentences and do not let go readability for briefness. You can maintain it succinct by phrasing sentences so that they provide more than lone rationale. The author can at this moment go straight to

shortening the outcome. Sum up the study, with the subsequent elements in any summary. Try to maintain the initial two items to no more than one ruling each.

- Reason of the study - theory, overall issue, purpose
- Fundamental goal
- To the point depiction of the research
- Consequences, including definite statistics - if the consequences are quantitative in nature, account quantitative data; results of any numerical analysis should be reported
- Significant conclusions or questions that track from the research(es)

Approach:

- Single section, and succinct
- As a outline of job done, it is always written in past tense
- A conceptual should situate on its own, and not submit to any other part of the paper such as a form or table
- Center on shortening results - bound background information to a verdict or two, if completely necessary
- What you account in an conceptual must be regular with what you reported in the manuscript
- Exact spelling, clearness of sentences and phrases, and appropriate reporting of quantities (proper units, important statistics) are just as significant in an abstract as they are anywhere else

Introduction:

The **Introduction** should "introduce" the manuscript. The reviewer should be presented with sufficient background information to be capable to comprehend and calculate the purpose of your study without having to submit to other works. The basis for the study should be offered. Give most important references but shun difficult to make a comprehensive appraisal of the topic. In the introduction, describe the problem visibly. If the problem is not acknowledged in a logical, reasonable way, the reviewer will have no attention in your result. Speak in common terms about techniques used to explain the problem, if needed, but do not present any particulars about the protocols here. Following approach can create a valuable beginning:

- Explain the value (significance) of the study
- Shield the model - why did you employ this particular system or method? What is its compensation? You strength remark on its appropriateness from a abstract point of vision as well as point out sensible reasons for using it.
- Present a justification. Status your particular theory (es) or aim(s), and describe the logic that led you to choose them.
- Very for a short time explain the tentative propose and how it skilled the declared objectives.

Approach:

- Use past tense except for when referring to recognized facts. After all, the manuscript will be submitted after the entire job is done.
- Sort out your thoughts; manufacture one key point with every section. If you make the four points listed above, you will need a least of four paragraphs.
- Present surroundings information only as desirable in order hold up a situation. The reviewer does not desire to read the whole thing you know about a topic.
- Shape the theory/purpose specifically - do not take a broad view.
- As always, give awareness to spelling, simplicity and correctness of sentences and phrases.

Procedures (Methods and Materials):

This part is supposed to be the easiest to carve if you have good skills. A sound written Procedures segment allows a capable scientist to replacement your results. Present precise information about your supplies. The suppliers and clarity of reagents can be helpful bits of information. Present methods in sequential order but linked methodologies can be grouped as a segment. Be concise when relating the protocols. Attempt for the least amount of information that would permit another capable scientist to spare your outcome but be cautious that vital information is integrated. The use of subheadings is suggested and ought to be synchronized with the results section. When a technique is used that has been well described in another object, mention the specific item describing a way but draw the basic

principle while stating the situation. The purpose is to text all particular resources and broad procedures, so that another person may use some or all of the methods in one more study or referee the scientific value of your work. It is not to be a step by step report of the whole thing you did, nor is a methods section a set of orders.

Materials:

- Explain materials individually only if the study is so complex that it saves liberty this way.
- Embrace particular materials, and any tools or provisions that are not frequently found in laboratories.
- Do not take in frequently found.
- If use of a definite type of tools.
- Materials may be reported in a part section or else they may be recognized along with your measures.

Methods:

- Report the method (not particulars of each process that engaged the same methodology)
- Describe the method entirely
- To be succinct, present methods under headings dedicated to specific dealings or groups of measures
- Simplify - details how procedures were completed not how they were exclusively performed on a particular day.
- If well known procedures were used, account the procedure by name, possibly with reference, and that's all.

Approach:

- It is embarrassed or not possible to use vigorous voice when documenting methods with no using first person, which would focus the reviewer's interest on the researcher rather than the job. As a result when script up the methods most authors use third person passive voice.
- Use standard style in this and in every other part of the paper - avoid familiar lists, and use full sentences.

What to keep away from

- Resources and methods are not a set of information.
- Skip all descriptive information and surroundings - save it for the argument.
- Leave out information that is immaterial to a third party.

Results:

The principle of a results segment is to present and demonstrate your conclusion. Create this part a entirely objective details of the outcome, and save all understanding for the discussion.

The page length of this segment is set by the sum and types of data to be reported. Carry on to be to the point, by means of statistics and tables, if suitable, to present consequences most efficiently.You must obviously differentiate material that would usually be incorporated in a study editorial from any unprocessed data or additional appendix matter that would not be available. In fact, such matter should not be submitted at all except requested by the instructor.

Content

- Sum up your conclusion in text and demonstrate them, if suitable, with figures and tables.
- In manuscript, explain each of your consequences, point the reader to remarks that are most appropriate.
- Present a background, such as by describing the question that was addressed by creation an exacting study.
- Explain results of control experiments and comprise remarks that are not accessible in a prescribed figure or table, if appropriate.
- Examine your data, then prepare the analyzed (transformed) data in the form of a figure (graph), table, or in manuscript form.

What to stay away from

- Do not discuss or infer your outcome, report surroundings information, or try to explain anything.
- Not at all, take in raw data or intermediate calculations in a research manuscript.

- Do not present the similar data more than once.
- Manuscript should complement any figures or tables, not duplicate the identical information.
- Never confuse figures with tables - there is a difference.

Approach

- As forever, use past tense when you submit to your results, and put the whole thing in a reasonable order.
- Put figures and tables, appropriately numbered, in order at the end of the report
- If you desire, you may place your figures and tables properly within the text of your results part.

Figures and tables

- If you put figures and tables at the end of the details, make certain that they are visibly distinguished from any attach appendix materials, such as raw facts
- Despite of position, each figure must be numbered one after the other and complete with subtitle
- In spite of position, each table must be titled, numbered one after the other and complete with heading
- All figure and table must be adequately complete that it could situate on its own, divide from text

Discussion:

The Discussion is expected the trickiest segment to write and describe. A lot of papers submitted for journal are discarded based on problems with the Discussion. There is no head of state for how long a argument should be. Position your understanding of the outcome visibly to lead the reviewer through your conclusions, and then finish the paper with a summing up of the implication of the study. The purpose here is to offer an understanding of your results and hold up for all of your conclusions, using facts from your research and generally accepted information, if suitable. The implication of result should be visibly described. Infer your data in the conversation in suitable depth. This means that when you clarify an observable fact you must explain mechanisms that may account for the observation. If your results vary from your prospect, make clear why that may have happened. If your results agree, then explain the theory that the proof supported. It is never suitable to just state that the data approved with prospect, and let it drop at that.

- Make a decision if each premise is supported, discarded, or if you cannot make a conclusion with assurance. Do not just dismiss a study or part of a study as "uncertain."
- Research papers are not acknowledged if the work is imperfect. Draw what conclusions you can based upon the results that you have, and take care of the study as a finished work
- You may propose future guidelines, such as how the experiment might be personalized to accomplish a new idea.
- Give details all of your remarks as much as possible, focus on mechanisms.
- Make a decision if the tentative design sufficiently addressed the theory, and whether or not it was correctly restricted.
- Try to present substitute explanations if sensible alternatives be present.
- One research will not counter an overall question, so maintain the large picture in mind, where do you go next? The best studies unlock new avenues of study. What questions remain?
- Recommendations for detailed papers will offer supplementary suggestions.

Approach:

- When you refer to information, differentiate data generated by your own studies from available information
- Submit to work done by specific persons (including you) in past tense.
- Submit to generally acknowledged facts and main beliefs in present tense.

Administration Rules Listed Before Submitting Your Research Paper to Global Journals Inc. (US)

Please carefully note down following rules and regulation before submitting your Research Paper to Global Journals Inc. (US):

Segment Draft and Final Research Paper: You have to strictly follow the template of research paper. If it is not done your paper may get rejected.

- The **major constraint** is that you must independently make all content, tables, graphs, and facts that are offered in the paper. You must write each part of the paper wholly on your own. The Peer-reviewers need to identify your own perceptive of the concepts in your own terms. NEVER extract straight from any foundation, and never rephrase someone else's analysis.

- Do not give permission to anyone else to "PROOFREAD" your manuscript.

- Methods to avoid Plagiarism is applied by us on every paper, if found guilty, you will be blacklisted by all of our collaborated research groups, your institution will be informed for this and strict legal actions will be taken immediately.)
- To guard yourself and others from possible illegal use please do not permit anyone right to use to your paper and files.

Criterion for Grading a Research Paper (Compilation) by Global Journals Inc. (US)

Please note that following table is only a Grading of "Paper Compilation" and not on "Performed/Stated Research" whose grading solely depends on Individual Assigned Peer Reviewer and Editorial Board Member. These can be available only on request and after decision of Paper. This report will be the property of Global Journals Inc. (US).

Topics	Grades		
	A-B	**C-D**	**E-F**
Abstract	Clear and concise with appropriate content, Correct format. 200 words or below	Unclear summary and no specific data, Incorrect form Above 200 words	No specific data with ambiguous information Above 250 words
Introduction	Containing all background details with clear goal and appropriate details, flow specification, no grammar and spelling mistake, well organized sentence and paragraph, reference cited	Unclear and confusing data, appropriate format, grammar and spelling errors with unorganized matter	Out of place depth and content, hazy format
Methods and Procedures	Clear and to the point with well arranged paragraph, precision and accuracy of facts and figures, well organized subheads	Difficult to comprehend with embarrassed text, too much explanation but completed	Incorrect and unorganized structure with hazy meaning
Result	Well organized, Clear and specific, Correct units with precision, correct data, well structuring of paragraph, no grammar and spelling mistake	Complete and embarrassed text, difficult to comprehend	Irregular format with wrong facts and figures
Discussion	Well organized, meaningful specification, sound conclusion, logical and concise explanation, highly structured paragraph reference cited	Wordy, unclear conclusion, spurious	Conclusion is not cited, unorganized, difficult to comprehend
References	Complete and correct format, well organized	Beside the point, Incomplete	Wrong format and structuring

INDEX

Q

R

S

T

U

V

W

save our planet

Global Journal of Computer Science and Technology

Visit us on the Web at www.GlobalJournals.org | www.ComputerResearch.org
or email us at helpdesk@globaljournals.org

ISSN 9754350