

# GLOBAL JOURNAL

OF COMPUTER SCIENCE AND TECHNOLOGY: C

---

## Software & Data Engineering

Proposed Adaptive Filter

Nomenclature and Contemporary

**Highlights**

Rule Generation Scheme

Text and Document Mining

Discovering Thoughts, Inventing Future

VOLUME 15

ISSUE 2

VERSION 1.0



GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY: C  
SOFTWARE & DATA ENGINEERING

---



GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY: C  
SOFTWARE & DATA ENGINEERING

---

VOLUME 15 ISSUE 2 (VER. 1.0)

OPEN ASSOCIATION OF RESEARCH SOCIETY

© Global Journal of Computer Science and Technology. 2015.

All rights reserved.

This is a special issue published in version 1.0 of "Global Journal of Computer Science and Technology" By Global Journals Inc.

All articles are open access articles distributed under "Global Journal of Computer Science and Technology"

Reading License, which permits restricted use. Entire contents are copyright by of "Global Journal of Computer Science and Technology" unless otherwise noted on specific articles.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without written permission.

The opinions and statements made in this book are those of the authors concerned. Ultraculture has not verified and neither confirms nor denies any of the foregoing and no warranty or fitness is implied.

Engage with the contents herein at your own risk.

The use of this journal, and the terms and conditions for our providing information, is governed by our Disclaimer, Terms and Conditions and Privacy Policy given on our website <http://globaljournals.us/terms-and-condition/menu-id-1463/>

By referring / using / reading / any type of association / referencing this journal, this signifies and you acknowledge that you have read them and that you accept and will be bound by the terms thereof.

All information, journals, this journal, activities undertaken, materials, services and our website, terms and conditions, privacy policy, and this journal is subject to change anytime without any prior notice.

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: *Open Association of Research Society*  
*Open Scientific Standards*

### *Publisher's Headquarters office*

**Global Journals Headquarters**  
301st Edgewater Place Suite, 100 Edgewater Dr.-Pl,  
**Wakefield MASSACHUSETTS, Pin: 01880,**  
**United States of America**

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

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

### *Offset Typesetting*

**Global Journals Incorporated**  
2nd, Lansdowne, Lansdowne Rd., Croydon-Surrey,  
**Pin: CR9 2ER, United Kingdom**

### *Packaging & Continental Dispatching*

**Global Journals**  
E-3130 Sudama Nagar, Near Gopur Square,  
**Indore, M.P., Pin:452009, India**

### *Find a correspondence nodal officer near you*

To find nodal officer of your country, please email us at [local@globaljournals.org](mailto:local@globaljournals.org)

### *eContacts*

Press Inquiries: [press@globaljournals.org](mailto:press@globaljournals.org)  
Investor Inquiries: [investors@globaljournals.org](mailto:investors@globaljournals.org)  
Technical Support: [technology@globaljournals.org](mailto:technology@globaljournals.org)  
Media & Releases: [media@globaljournals.org](mailto: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)

INTEGRATED EDITORIAL BOARD  
(COMPUTER SCIENCE, ENGINEERING, MEDICAL, MANAGEMENT, NATURAL  
SCIENCE, SOCIAL SCIENCE)

**John A. Hamilton, "Drew" Jr.,**  
Ph.D., Professor, Management  
Computer Science and Software  
Engineering  
Director, Information Assurance  
Laboratory  
Auburn University

**Dr. Henry Hexmoor**  
IEEE senior member since 2004  
Ph.D. Computer Science, University at  
Buffalo  
Department of Computer Science  
Southern Illinois University at Carbondale

**Dr. Osman Balci, Professor**  
Department of Computer Science  
Virginia Tech, Virginia University  
Ph.D. and M.S. Syracuse University,  
Syracuse, New York  
M.S. and B.S. Bogazici University,  
Istanbul, Turkey

**Yogita Bajpai**  
M.Sc. (Computer Science), FICCT  
U.S.A. Email:  
yogita@computerresearch.org

**Dr. T. David A. Forbes**  
Associate Professor and Range  
Nutritionist  
Ph.D. Edinburgh University - Animal  
Nutrition  
M.S. Aberdeen University - Animal  
Nutrition  
B.A. University of Dublin- Zoology

**Dr. Wenying Feng**  
Professor, Department of Computing &  
Information Systems  
Department of Mathematics  
Trent University, Peterborough,  
ON Canada K9J 7B8

**Dr. Thomas Wischgoll**  
Computer Science and Engineering,  
Wright State University, Dayton, Ohio  
B.S., M.S., Ph.D.  
(University of Kaiserslautern)

**Dr. Abdurrahman Arslanyilmaz**  
Computer Science & Information Systems  
Department  
Youngstown State University  
Ph.D., Texas A&M University  
University of Missouri, Columbia  
Gazi University, Turkey

**Dr. Xiaohong He**  
Professor of International Business  
University of Quinnipiac  
BS, Jilin Institute of Technology; MA, MS,  
PhD,. (University of Texas-Dallas)

**Burcin Becerik-Gerber**  
University of Southern California  
Ph.D. in Civil Engineering  
DDes from Harvard University  
M.S. from University of California, Berkeley  
& Istanbul University

**Dr. Bart Lambrecht**

Director of Research in Accounting and Finance  
Professor 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 Regina  
Ph.D., M.Sc. in Mathematics  
B.A. (Honors) in Mathematics  
University of Windsor

**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., Eötvös Loránd University  
Postdoctoral Training,  
New York University

**Dr. Söhnke M. Bartram**

Department of Accounting and Finance  
Lancaster University Management School  
Ph.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  
Neuroscience  
Northwestern 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

## PRESIDENT EDITOR (HON.)

---

### **Dr. George Perry, (Neuroscientist)**

Dean and Professor, College of Sciences

Denham Harman Research Award (American Aging Association)

ISI Highly Cited Researcher, Iberoamerican Molecular Biology Organization

AAAS Fellow, Correspondent Member of Spanish Royal Academy of Sciences

University of Texas at San Antonio

Postdoctoral Fellow (Department of Cell Biology)

Baylor College of Medicine

Houston, Texas, United States

## CHIEF AUTHOR (HON.)

---

### **Dr. R.K. Dixit**

M.Sc., Ph.D., FICCT

Chief Author, India

Email: [authorind@computerresearch.org](mailto:authorind@computerresearch.org)

## DEAN & EDITOR-IN-CHIEF (HON.)

---

### **Vivek Dubey(HON.)**

MS (Industrial Engineering),

MS (Mechanical Engineering)

University of Wisconsin, FICCT

Editor-in-Chief, USA

[editorusa@computerresearch.org](mailto:editorusa@computerresearch.org)

### **Sangita Dixit**

M.Sc., FICCT

Dean & Chancellor (Asia Pacific)

[deanind@computerresearch.org](mailto:deanind@computerresearch.org)

### **Suyash Dixit**

(B.E., Computer Science Engineering), FICCTT

President, Web Administration and

Development , CEO at IOSRD

COO at GAOR & OSS

### **Er. Suyog Dixit**

(M. Tech), BE (HONS. in CSE), FICCT

SAP Certified Consultant

CEO at IOSRD, GAOR & OSS

Technical Dean, Global Journals Inc. (US)

Website: [www.suyogdixit.com](http://www.suyogdixit.com)

Email: [suyog@suyogdixit.com](mailto:suyog@suyogdixit.com)

### **Pritesh Rajvaidya**

(MS) Computer Science Department

California State University

BE (Computer Science), FICCT

Technical Dean, USA

Email: [pritesh@computerresearch.org](mailto:pritesh@computerresearch.org)

### **Luis Galárraga**

J!Research Project Leader

Saarbrücken, Germany



## CONTENTS OF THE ISSUE

---

- i. Copyright Notice
  - ii. Editorial Board Members
  - iii. Chief Author and Dean
  - iv. Contents of the Issue
- 
1. P<sup>2</sup>DM-RGCD: PPDM Centric Classification rule Generation Scheme. *1-9*
  2. Removal of Power Line Interference From Electrocardiograph (ECG) Using Proposed Adaptive Filter Algorithm. *11-14*
  3. Nomenclature and Contemporary Affirmation of the Unsupervised Learning in Text and Document Mining. *15-21*
  4. Changing Scenario of Testing Paradigms Using DevOps - A Comparative Study with Classical Models. *23-27*
- 
- v. Fellows and Auxiliary Memberships
  - vi. Process of Submission of Research Paper
  - vii. Preferred Author Guidelines
  - viii. Index



# P<sup>2</sup>DM-RGCD: PPDM Centric Classification rule Generation Scheme

By S Kumara Swamy, Manjula S H, K R Venugopal & L M Patnaik

*UVCE, Bangalore University, India*

**Abstract-** In present day applications the approach of data mining and associated privacy preservation plays a significant role for ensuring optimal mining function. The approach of privacy preserving data mining (PPDM) emphasizes on ensuring security of private information of the participants. On the contrary majority of present mining applications employ the vertically partitioned data for mining utilities. In such scenario when the overall rule is divided among participants, some of the parties remain with fewer rules sets and thus the classification accuracy achieved by them always remain questionable. On the other hand, the consideration of private information associated with any part will violate the approach of PPDM. Therefore, in order to eliminate such situations and to provide a facility of rule regeneration in this paper, a highly robust and efficient rule regeneration scheme has been proposed ensures optimal classification accuracy without using any critical user information for rule generation. The proposed system developed a rule generation function called cumulative dot product (P<sup>2</sup>DM-RGCD) rule regeneration scheme. The developed algorithm generates two possible optimal rule generation and update functions based on cumulative updates and dot product. The proposed system has exhibited optimal response in terms of higher classification accuracy, minimum information loss and optimal training efficiency.

**Keywords:** *data mining, privacy preserving, vertical portioning, rule regeneration.*

**GJCST-C Classification :** *H.3.4*



*Strictly as per the compliance and regulations of:*



# P<sup>2</sup>DM-RGCD: PPDM Centric Classification rule Generation Scheme

S Kumara Swamy<sup>α</sup>, Manjula S H<sup>σ</sup>, K R Venugopal<sup>ρ</sup> & L M Patnaik<sup>ω</sup>

**Abstract-** In present day applications the approach of data mining and associated privacy preservation plays a significant role for ensuring optimal mining function. The approach of privacy preserving data mining (PPDM) emphasizes on ensuring security of private information of the participants. On the contrary majority of present mining applications employ the vertically partitioned data for mining utilities. In such scenario when the overall rule is divided among participants, some of the parties remain with fewer rules sets and thus the classification accuracy achieved by them always remain questionable. On the other hand, the consideration of private information associated with any part will violate the approach of PPDM. Therefore, in order to eliminate such situations and to provide a facility of rule regeneration in this paper, a highly robust and efficient rule regeneration scheme has been proposed ensures optimal classification accuracy without using any critical user information for rule generation. The proposed system developed a rule generation function called cumulative dot product (P<sup>2</sup>DM-RGCD) rule regeneration scheme. The developed algorithm generates two possible optimal rule generation and update functions based on cumulative updates and dot product. The proposed system has exhibited optimal response in terms of higher classification accuracy, minimum information loss and optimal training efficiency.

**Keywords:** data mining, privacy preserving, vertical partitioning, rule regeneration.

## I. INTRODUCTION

In present day scenario the data mining techniques are playing very significant role for ensuring optimal data exploration, classification and further decision support systems (DSS). In numerous applications the process of data mining is having great significance such as search engines and DSS mechanisms for business houses, organizations and government agencies etc. On the other hand due to multi-party computation or communication scenario, the requirement of a robust privacy factor is realized. A number of researches are going on to ensure private data security in secure multiparty computation (SMC) scenario based mining facility. The newly proposed paradigm called Privacy preservation in data mining (PPDM) is one of the growing research sector where a number of approaches have been proposed and optimized for optimal and secure mining process. In order to achieve an optimal

and secure mining facility, data distribution approaches such as vertical partitioning and horizontal data partitioning has been advocated. The systems based on vertically partitioned data are emerging due to its robust function and classification accuracy. On the other hand based on association rule mining a number of systems have been developed. In our previous research [1][2][3] we have already implemented numerous noble schemes to optimize data classification and performance efficiency and a robust privacy preserving data mining scheme using commutative RSA scheme. These all system has in fact exhibited optimal performance for classification efficiency and effective mining function. But taking into consideration of a scenario, where in vertically partitioned data the rules generated have to be divided among encompassing participants, there could be a possibility that some of the parties might have fewer rules.

When certain party possesses low rules count, the classification accuracy based on those confined rules might give lower accuracy. Therefore to ensure optimal classification accuracy and efficiency rules are required to be increased with enhanced information and classification attributes. On the contrary, in privacy preserving data mining (PPDM) scenario, no other party will like to share its critical, private information with other and if it takes place the PPDM itself will be violated. Therefore in such circumstances, the implementation of such approach which can ensure rule enhancement or rule regeneration without retrieving critical information of other participant will be required. In order to achieve this goal, here in this research paper, we have proposed a highly robust and efficient system model for rule regeneration which considers only some of the numerical attributes for rule regeneration and operates with two mathematical and logical operators. In this paper a rule regeneration scheme called cumulative dot product (P<sup>2</sup>DM-RGCD) has been proposed. The proposed scheme individually generates two distinct functions for rule regeneration on the basis of cumulative and dot product based rule updates. The, overall functions and rule regeneration schemes have been developed employing only some of the numerical attributes associated with other parties so as to perform rule regeneration. The considered numerical attributes even doesn't disclose the private or critical information related to parties. Thus, the proposed approaches of rule regeneration not only ensure the preservation of

*Author α σ ρ* : Department of Computer Science and Engineering University Visvesvaraya College of Engineering, Bangalore University, Bangalore. e-mail: kumar.aruna@gmail.com

*Author ω* : Indian Institute of Science, Bangalore, India.

privacy but also make the mining system optimized in terms of higher classification accuracy and efficiency. The proposed system has exhibited optimal response in terms of higher classification accuracy, minimum information loss and optimal training efficiency.

#### a) *Motivation*

Now days most of the researches are done for privacy preservation in mining but very few have made effort to ensure optimal performance with PPDM as most schemes employ the approach of vertically partitioned data for classification and in case of SMC scenario, the requirement of privacy preserving data mining is also inevitable. The situation also becomes complex in the scenario, where the overall rules are required to be split amongst allied participants. In this case, some of the parties remained back with low rule counts and therefore their classification accuracy is always under suspicion. Therefore, considering this requirement the development of certain optimized rule enhancement or even rule regeneration scheme can be a potential solution. A model based on PPDM is needed where an algorithm will regenerate the rules based on some numerical attributes using some operators which neither use critical information of other users nor violates PPDM objective. These all motivate us to develop a robust rule regeneration scheme that can exhibit rule regeneration without causing any violation in PPDM. The proposed cumulative dot product (P<sup>2</sup>DM-RGCD) scheme can deliver these all expectations as it doesn't employ critical information of user, rather considers only some numerical attributes have not much information. The mathematical function (dot product and cumulative updates) makes the system more robust in function.

#### b) *Contribution*

The proposed rule regeneration approach, cumulative dot product (P<sup>2</sup>DM-RGCD) scheme possesses potential to optimize rule regeneration on the basis of some numerical attributes associated with any participants in MPC scenario. The proposed system performs well using two possible functions, in terms of cumulative rule generation updates and dot product based rule generation. Such combinations emerge out with enhanced rule generation efficiency, classification accuracy, minimum information loss and higher training efficiency. The developed system has been tested with varied datasets of varying sample size and the results obtained has exhibited that the proposed system can play a significant role for real time mining applications.

#### c) *Organization*

The remaining manuscript has been classified into certain sections where Section II represents related Work which are followed by research background. In Section III the proposed system has been discussed which is followed by results and analysis in Section IV. Section V presents conclusion. The references used are given at the last of presented manuscript.

## II. RELATED WORK

A number of researches have been done for PPDM oriented rule generation and performance optimization. Some of the work carried for PPDM and rule generation based mining enhancements are as follows:

Dehzangi, O. [4] Advocated on the application of fuzzy rule based systems and discusses the limitations in terms of rule-base generation and stated that in case of higher dimensional issues, not every possible rule can be generation correspondence with entire antecedent combinations. Ultimately authors proposed a rule generation approach using data mining and focused their system to accomplish rule-based generation with varied length. In [2] M.W. Kim et al. developed an effective fuzzy based rule generation scheme using fuzzy decision tree data mining approach and they combined the clarity of rules generated on the basis of decision tree approaches like ID3 and C4.5 enriched with presentative ability of fuzzy sets that facilitated better classification for varied patterns associated with non axis-parallel decision boundaries that is in fact intricate for implementation employing attribute-based classification scheme. Sabu, M.K. et al., [6] analyzed a recent scheme called Rough Set Theory (RST) and stated it as a system with ambiguity and insecurity. In fact RST is significant for various applications but cannot incorporate association rules that plays significant rule for data mining while ensuring association among varied attributes. To eliminate such limitation the author advocated a rough set based scheme for rule generation using an incoherent information model comprising preprocessed data and used LEM2 algorithm to perform rule generation. Ji Dan et al., [7] presented a data mining scheme called CA to enhance CURE and C4.5 and uses principle component analysis (PCA), parallel processing and grid partitioning to perform better feature and scale reduction for huge datasets. Trinčá, D. et al., [8] emphasized on rule mining based PPDM and proposed an algebraic and recursive system based on two party protocols and focussed on collusion free mining still. In our last paper [9] we accomplished data mining while incorporating multiple parties and performed mining on vertically partitioned data and proposed a scheme called Key Distribution-Less Privacy Preserving Data Mining (KDLPPDM). To ensure security they employed Commutative RSA an advanced cryptosystem. Tran, D.H. et al., [10] proposed CRYPPAR scheme that facilitates a robust framework for privacy preserving association rule mining based on cryptosystem schemes. The authors employed secure scalar product algorithms for exhibiting efficient data mining with enhanced accuracy. Modi, C.N. et al., [11] proposed a noble heuristic scheme called decrease support of R.H.S. item of rule clusters that facilitates privacy for

perceptive rules at definite level while assuring optimal quality or mining efficiency for datasets. They performed clustering on the sensitive association rules based on defined conditions and perform rule hiding by means of some modifications. This is the matter of fact the some of the existing approaches have illustrated better results but unfortunately, no emphasis has been made of system optimization using PPDM approach with rule regeneration without exploring critical information of associated parties. Some works either focuses on PPDM or classification accuracy, but for robust applications, the duo are needed to be enriched together.

### III. BACKGROUND WORK

PPDM is one of the recent and most emerging technologies for data mining filed. This technology facilitates a novel framework for performing data extraction and classification with ensured security and preservation among various or multiple parties. In [1] a noble system model for PPDM has been developed for vertically partitioned data and authors have employed a robust cryptosystem to ensure data security in SMC environment. Commutative RSA scheme has been used for privacy preservation. Similarly in [9] the emphasis was made on classification accuracy. In uniqueness of this work was that this algorithm didn't employ any private data and in spite it came up with better association rule mining. This system came up with better efficiency in terms of rule generation, overhead minimization and classification efficiency.

### IV. PROPOSED SYSTEM

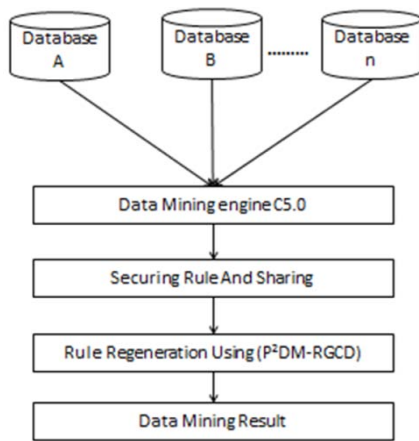


Figure 1 : PPDM-RGCD System Architecture

Taking into consideration of a highly robust and efficient system for privacy preserving data mining that employs multiple parties with vertically partitioned data and in which the rules associated with certain party defines the accuracy and performance, here in this research paper, an optimal solution of rule regeneration and performance optimization has been developed. The Prime objective of this research work is to develop a rule

regeneration scheme without employing any critical information of allied parties. In this paper a noble cumulative dot product (P<sup>2</sup>DM-RGCD) scheme has been proposed for rule regeneration with the party having fewer rules, without retrieving any critical information of other parties can exhibit higher rule generation, resulting into better classification accuracy and efficiency. The proposed scheme has been discussed in this section.

#### a) PPDM Oriented Rule Regeneration Scheme

As in vertical partitioned data the overall rules are shared among the participating parties and thus it raises the probability where the rules available with certain party could be very few and on that basis the classification accuracy could not be optimal. On the other hand, sharing the private information about other parties to retrieve better classification for certain party is violation for privacy preservation. Therefore in such situation, in this paper rule regenerates have been done based on certain numerical attributes. In order to accomplish an optimized rule regeneration approach for PPDM applications, a classifier having mapping function  $g(a)$  with  $a \in A$  into labels  $b \in \{-1, 1\}$ , are taken into consideration and is given by

$$g(a) = \text{sgn}[p(a)] \tag{1}$$

Where  $p : A \rightarrow C$  states a real-valued predictor. The optimal predictor  $p(a)$  represents the reduction factor for the classification problem. Mathematically,

$$C(p) = D_{A,B}\{M[bp(a)]\} \approx \sum_i M[b_i p(a_i)] \tag{2}$$

In above equation  $D(\cdot)$  represents the function for loss reflecting the upper bound as the error rate. The process of rule regeneration approach performs learning and approximating the best suitable predictor in terms of a linear combination of generic predictors  $g_n : A \rightarrow C$ . In fact this is stated as the party possessing lower classification rules and information to exhibit optimal classification. Thus, the reduction of error minimization function can be updated as

$$p(a) = \sum_n x_n h_n(a) \quad h \in H \tag{3}$$

Where  $H = \{h_1, \dots, h_m\}$  refers for the combination or cluster of those all parties which don't have enough rules or information to perform classification. In order to achieve formulation consistency or regularity here we have considered that  $h(a) = 0$  and  $h(a) = 1$  are the parts of cluster or group of parties ( $H$ ) having less rules. The process of rules regeneration performs learning of the linear combination by employing certain successive descent approach in functional space, in reference to the enhancement issues,

$$\begin{cases} \min_{p(a)} C(p) \\ s.t. tp(a) \in \Phi_h \end{cases} \quad (4)$$

In the above presented expression, the variable  $H$  represents the cluster of parties possessing fewer rules for classification. Here it must be taken into consideration that  $H$  represents a convex set and in case the parameter  $C(\cdot)$  depicts certain convex function then in that situation the enhancement issues for equation (3) will always be a convex. Now, taking into consideration of the first and second derivatives of the loss function as employed with initial problem (Equation 2) is stated by

$$M' = \frac{\partial M(q)}{\partial q} \text{ and } M'' = \frac{\partial^2 M(q)}{\partial q^2} \quad (5)$$

Here it is assumed that, performing  $k$  iterations the value of the optimal predictor is given by a function  $p^n(a)$ . Now employing Taylor series expansion for  $C(p^n + h)$  towards  $p^n$ , the resulting first and second derivatives can be retrieved by the following approach.

$$\delta C = \left. \frac{\delta C(p^n + \gamma h)}{\partial \gamma} \right|_{\gamma=0} \quad (6)$$

It provides the estimation of variations in  $C$  at instant  $p^n$  towards  $g$ . To make it simple, in many cases the argument  $a$  has been omitted all through the presented manuscript. The variations or warp of  $C[p^n]$  towards  $g$  can be presented in terms of its second order derivative. Mathematically the curvature can be defined as

$$\delta^2 C(p^n; h) = \left. \frac{\delta^2 C(p^n + \gamma h)}{\partial \gamma^2} \right|_{\gamma=0} \quad (7)$$

Now, taking into consideration of these variables, the approximation of  $C$  can be accomplished using Taylor series expansion in the region of since  $C[p^n]$ , given below, solely depend on the illustrative value  $A$  through the training events  $a_i$  and there doesn't exist any loss while mapping the value of defined functions  $p^n(a)$  into certain definite vector form  $Q \in C^e$  of  $[p^n(a_1), \dots, p^n(a_e)]$ , where  $n$  states for the size of the training set.

$$\begin{aligned} C(p^n + \beta g) &= C(p^n) + \beta \delta C(p^n; h) \\ &+ \frac{\beta^2}{2} \delta^2 C(p^n; h) \\ &+ E(\beta^2) \end{aligned} \quad (8)$$

In the defined vector  $Q$  the dot product presentation can be given by

$$\langle h_1, h_2 \rangle = \sum_i h_1(a_i), h_2(a_i) \quad (9)$$

Without causing any loss in generality, it is assumed that those parties who have fewer rule for classification ( $h \in H$ ), are in general processed for

normalization to get  $\langle h, h \rangle = 1$ . Now, considering a function called unitary indication function given by  $K(a) = 1$  in case the variable  $a$  holds otherwise it possesses zero and the functional gradient of  $C$  is given by certain vectorized entities given by

$$\begin{aligned} \nabla C_{(p^n)}(a_i) &= \left. \frac{\partial}{\partial \gamma} C[p^n + \gamma K(a = a_i)] \right|_{\gamma=0} \\ &= \left. \frac{\partial M[b_i, p^n(a_i) + \gamma]}{\partial \gamma} \right|_{\gamma=0} \end{aligned} \quad (11)$$

Meanwhile, the second order gradient vector is given as Hessian in the form of a matrix, mathematically it is presented as

$$\begin{aligned} &\nabla_{C_{(p^n)}}^2(a_i, a_j) \\ &= \left. \frac{\partial^2 C(p^n + \gamma_1 K(a = a_i) + \gamma_2 K(a = a_j))}{\partial \gamma_1 \partial \gamma_2} \right|_{\gamma_1, \gamma_2=0} \quad (12) \\ &= \begin{cases} \frac{\partial^2}{\partial \xi \gamma^2} M[b_i p^n(a_i) + \gamma] & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

To make it simple, it is presented further as

$$\begin{aligned} &\nabla_{C_{(p^n)}}^2(a_i, a_j) \text{ as } \nabla_{C_{(p^n)}}^2(a_i) \cdot C \\ &\nabla_{C_{(p^n)}}^2(a_i) \geq 0, \forall p^n, \forall a_i \end{aligned} \quad (13)$$

Now taking into consideration of the convex problem and parties to be enriched with rule regeneration, its projects comes out to be

$$\begin{aligned} \delta C(p^n; h) &= \left. \frac{\partial}{\partial \gamma} \sum_j M[b_j, p^n(a_j) + \gamma h(a_j)] \right|_{\gamma=0} \\ &= \sum_j h(a_j) \left. \frac{\partial}{\partial \gamma} M[b_i, p^n(a_j) + \sigma] \right|_{\sigma=0} \\ &= \langle \nabla_{C_{(p^n)}}, h \rangle \end{aligned} \quad (14)$$

Ultimately, the first order derive has been obtained as following equation (Equation 15) and similarly the second derivate has been obtained in terms of Equation (Equation 17).

$$\sum_i b_i h(a_i) M'[b_i p^n(a_i)] \quad (15)$$

$$\delta^2 C(p^n; \gamma h) = \left. \frac{\delta^2 C(p^n + \beta h)}{\partial \gamma^2} \right|_{\gamma=0} \quad (16)$$

$$\sum_i g^2(x_i) L''[y_i f^k(x_i)] \quad (17)$$

Considering the above derived expressions, it can be found that the learner party is required to be enriched with the predictor or rule generator at its  $n + 1$  iteration and it is given by

$$h^* = \underset{h \in H}{\operatorname{arg\,min}} \delta C(p^n; h) \tag{18}$$

In case of the consideration of the gradient factor, it can also be represented by

$$h^* = \underset{h \in H}{\operatorname{arg\,max}} \frac{[\delta C(p^n; h)]^2}{\delta^2 C(p^n; h)} \tag{19}$$

With a bet rule generator or party seeking rule regenerator  $h^*$ , the best possible step size can be presented by

$$x^* = \underset{x \in C}{\operatorname{arg\,min}} C(p^{nk}; xh^*) \tag{20}$$

Similarly, the values evaluated by predictor is updated for  $n + 1$  iteration events and it is given by

$$p^{n+1}(a) = p^n(a) + x^*h^*(a). \tag{21}$$

Thus, the ultimate predictor for rule regeneration will function as a linear combination of those parties who have fewer rules for classification.

a) *Rule Regeneration: An Optimal Blend for new Features or Rules*

This is the matter of fact that the rule regeneration selects some helpful features to perform classification; the genuine combination of parties seeking rule regeneration might not be so enriched for capturing every associated attributes of information to perform discrimination. As for illustration, it becomes necessary to employ certain conjunctions of the features for capturing few of the dimensions and in such conditions the linear combination becomes ineffective where even rule generation is not optimal solution to accomplish better classification. In this paper a mechanism of rule regeneration has been developed to ensure optimal privacy preservation and effective classification in terms of accuracy and optimal mining results. The proposed *cumulative dot product P<sup>2</sup>DM\_RG-CD* scheme has been discussed as follows.

b) *Cumulative Dot Product (P2dm\_Rg-Cd) Based Rule Regeneration*

Consider cluster of combination of parties having lower rules count  $G$ , the proposed noble scheme of rule generation called cumulative dot product (P<sup>2</sup>DM-RGCD) based rule generation emphasizes its function for solving the following problem

$$\begin{cases} \min_{p(a)C(p)} \\ \text{s.t. } p(a) \in \alpha_G^{CD} \end{cases} \tag{22}$$

Where  $\alpha_H^{CD}$  represents the combination of all comprising set of achievable linear combinations of the dot product of parties, mathematically given as

$$\alpha_H^{CD} = \left\{ g(a) \mid g(a) = \sum_j \prod_l h_j k(a), h_j k \in H \right\} \tag{23}$$

Here it can be found that  $\alpha_H^{CD}$  states for a convex combination and therefore for any similar functions  $C(\cdot)$ , the optimization issues will be convex. Now, considering Taylor boost scheme as a comparative model, assume that after  $n$  iterations the predictor possesses  $m$  terms given by

$$p^n(a) \sum_{j=1}^m r_j^n(a) \tag{24}$$

The every such presentation will represent an unitary learner party and will be given by,

$$r_j^n(a) \prod_{k=1}^{t_j} h_{j,k}(a), h_{j,k}(a) \in H \tag{25}$$

At certain iteration  $n + 1$  it is feasible to enhance  $p^n(a)$  possessing dual updates given by cumulative and dot product. A brief of the considered paradigms have been given as follows:

- *Cumulative Update:* In case of cumulative update paradigm it is considered that selecting or joining a learner party to the predictor will be like

$$p^{n+1}(a) = p^n(a) + h(a). \tag{26}$$

Here, the updates are done on the basis of rules regenerated with first and second derivatives factors given by  $\delta C(p^n; h)$  and  $\delta^2 C(p^n; h)$  respectively. And the optimal party  $h_0^*$ , can be achieved based on the selection of gradient descent approach. Here in terms of optimal step size  $x_0^*$ , the newly generated rules or predictor is found to be with risk factor

$$\hat{C}_0 = C(p^n x_0^* h_0^*) \tag{27}$$

- *Dot product rule update:* In case of *P<sup>2</sup>DM\_RG\_DC*, one of the available terms is processed for multiplication using a newer party given by

$$r_s^{n+1}(a) = r_s^n(x) \times g(a). \tag{28}$$

It can also be given by

$$p_s^{n+1}(a) = r_s^n(a)h(a) + \sum_{j \neq s} r_j^n(a) \tag{29}$$

$$= p^n(a) - r_s^n(a) + r_s^n(a)h(a) \tag{30}$$

$$F_s^n(a) + r_s^n(a)h(a) \tag{31}$$

$$\text{With } F_s^n(a) = p^n(a) - r_s^n(a). \tag{32}$$

Now, taking into consideration of the above mentioned expressions a Taylor series expansion of  $C(p^{n+1})$  can be retrieved in the region of functional  $F_s^n(a)$ , and the first and second order variations for the risk factor wrt a dot product cum cumulative update of the  $s^{th}$  term in  $p^n(a)$  is given by

$$\delta C(p^n; h, s) = \left. \frac{\partial C[F_s^n + \omega r_s^n h]}{\partial \omega} \right|_{\omega=0} \tag{33}$$

$$= \sum_i b_i h(a_i) r_s^n(a_i) ]^2 M'' [b_i F_s^n(a_i)] \tag{34}$$

$$\delta^2 C(p^n; h, s) = \frac{\delta^2 C[F_s^n + \omega r_s^n h]}{\partial \omega^2} \Big|_{\omega=0} \quad (35)$$

$$= \sum_i [h(a_i) r_s^n(a_i)]^2 M'' [b_i F_s^n(a_i)] \quad (36)$$

**Pseudo Algo: Cumulative dot product (P<sup>2</sup>DM-RGCD) Based Rule Regeneration**

**Input:** Sets for data training  $S_t$ , parties with lower rules  $H = \{h_1, \dots, h_m\}$ , Iteration counts  $Z$  and a loss function  $M(\cdot)$ .  
**Initialization:** Select  $n = 0, m = 0, r_m^n(a) = 0$  and  $p^n(a) = 0$   
**while**  $n < Z$  **do**  
 Estimate optimal Cumulative update  $x_0^* h_0^*$   
 Select  $\hat{C}_0 = C(p^n + x_0^* h_0^*)$   
 Initialize look  
**for**  $s = 1$  to  $m$  **do**  
 Estimate optimal update for  $s^{th}$  dot product term,  $x_s^* h_s^*$   
 Update  $\hat{C}_0 = C(p^n - r_s^n) + r_s^n x_s^* h_s^*$   
**end for**  
**Select**  $s^* = \arg \min_s \hat{C}_s, s = 0, \dots, m$ ,  
**if**  $s^* = 0$  **then**  
 $r_{m+1}^{n+1} = x_0^* h_0^*$   
 $m = m + 1$   
**else**  
 $r_{s^*}^{n+1} = r_{s^*}^n, s \neq s^*$   
 $p^{n+1}(a) = \sum_{s=1}^m r_s^{n+1}(a)$   
 $n = n + 1$   
**end while**  
**Output:** Rule generated:  $\text{sign}[p^Z(a)]$

The optimal party seeking rule regeneration given by  $g_r^*$  is retrieved with its optimal step size and it can be given as

$$x_s^* = \arg \min_{x \in C} C(F_s^n + x h_s^*) \quad (37)$$

Still, the updated rule generator or predictor does possess the risk factor given by

$$\hat{C}_s = C(F_s^n x_s^* h_s^*) \quad (38)$$

The proposed and developed algorithm of cumulative dot product (P<sup>2</sup>DM-RGCD) based rule regeneration provides optimal new rules for those parties who don't have sufficient rules for classification due to vertically partitioned data and divided rules sets among other parties. Thus, implementing the above mentioned paradigms cumulative update and dot product cum cumulative update has exhibited higher rule generation without extracting any critical information associated with the other parties in the application scenario and thus it also preserves the privacy of participant. The enriched rules or regenerated rules make the system highly robust for optimal classification.

## V. RESULTS AND ANALYSIS

In a specific situation of SMC based PPDM where a number of participants do exhibit data mining

without any disclosure of its private data or information with vertically partitioned data the splitting of rules for classification might cause a situation where some of the participants will have fewer numbers of rules that could result into inaccuracy, error prone and inefficient classification. In such cases even the other parties don't wish to share its information. In this paper a robust rule regeneration technique has been developed that exhibits rule regeneration using a noble scheme called cumulative dot product (P<sup>2</sup>DM-RGCD) without using significant information of other participants. C# and C++ programming languages was used for development. The model was implemented with GCC compiler on Linux platform and the system effectiveness has been analyzed in terms of its learning accuracy, testing accuracy, information loss etc. In order to exhibit the performance analysis with varied datasets or data count, various data samples like breast cancer data, diabetes datasets, satellite datasets etc have been considered. The results have been analyzed in terms of its specificity Vs sensitivity, performance rate, higher accuracy and minimum computational overheads information loss, classification accuracy and many more. Following figures represent the receiver operating characteristics (ROC) analysis for the developed research model. The results obtained in this paper have been compared with our previous work [1][2][3]. Figure 1-3, illustrates the performance of the proposed system with employed breast cancer data of varied size. Here it can be found that the proposed system response is better as compared to existing vertically partitioned mining model with PPDM.

a) Performance Analysis for Breast Cancer datasets

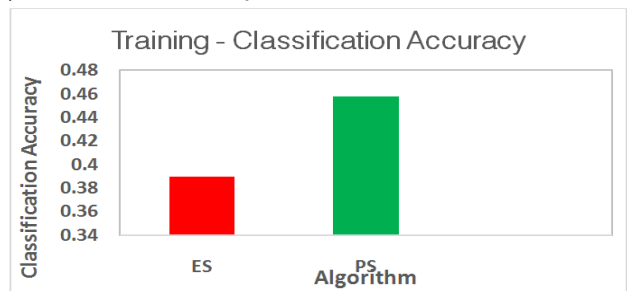


Figure 2 : Performance analysis for classification accuracy for Breast Cancer data

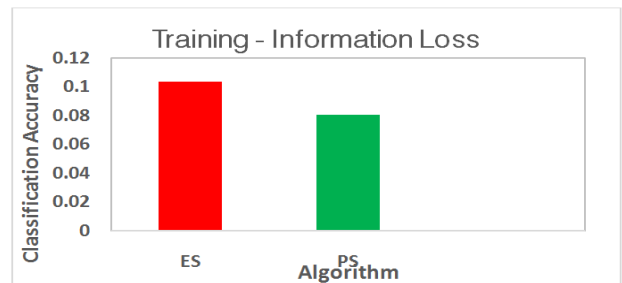


Figure 3 : Analysis for Information loss for Breast Cancer datasets



From Figure 2, Figure 4 and Figure 6, it can be found that the proposed system facilitates minimum information loss. The reason behind this achievement is that the proposed system does not employ the critical information associated with any participants. In order to exhibit rule regeneration, our proposed system has just employed some of the numeric values or parameters on basis of which processing with proposed cumulative dot product (P<sup>2</sup>DM-RGCD)scheme, the classification has been accomplished. Thus, the least utilization of critical information makes this system capable of delivering higher classification accuracy and performance (Figure1, Figure 4, and Figure 7) without causing much information loss as compared to existing systems.

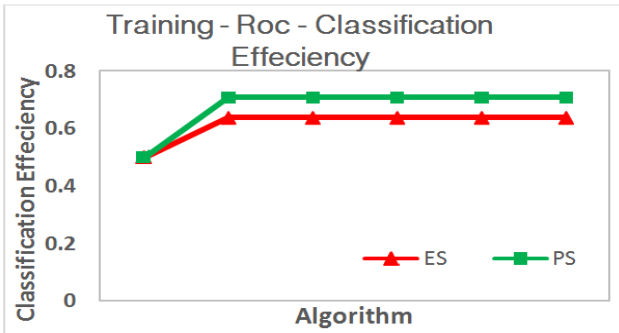


Figure 4 : ROC Analysis for Classification accuracy for Breast Cancer datasets

b) Performance Analysis for Diabetes datasets

The results obtained for diabetes datasets are as follows:

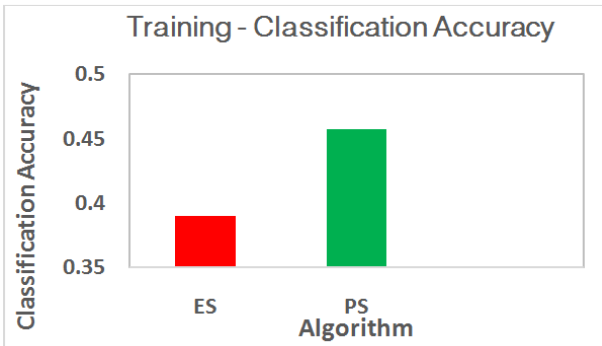


Figure 5 : Performance analysis for classification accuracy for Diabetic dataset

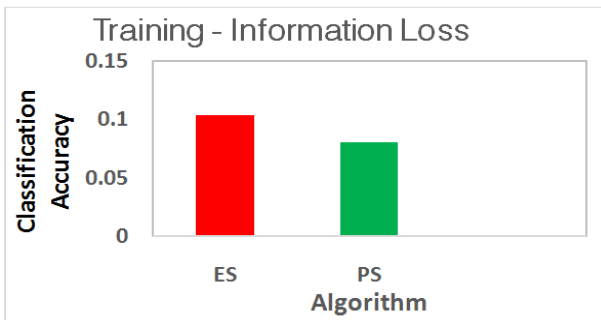


Figure 6 : Analysis for Information loss for Diabetes datasets

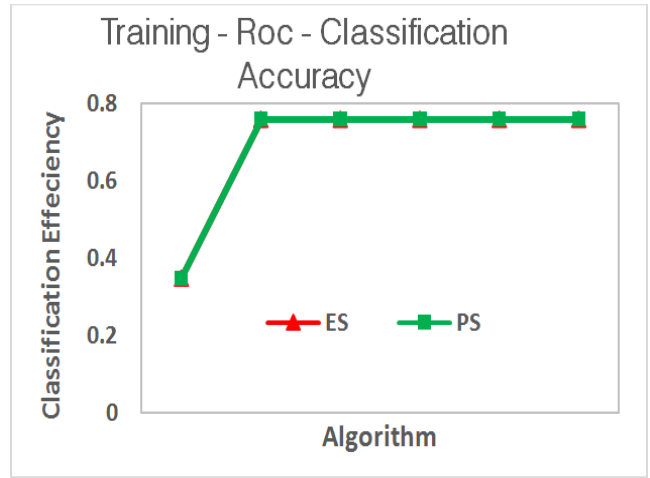


Figure 7 : ROC Analysis for Classification accuracy for Diabetes datasets

c) Performance Analysis for Satellite datasets

The results obtained for diabetes datasets are as follows:

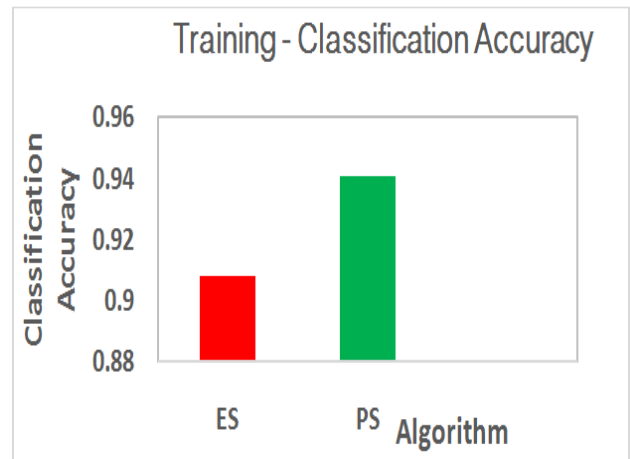


Figure 8 : Performance analysis for classification accuracy for Satellite datasets

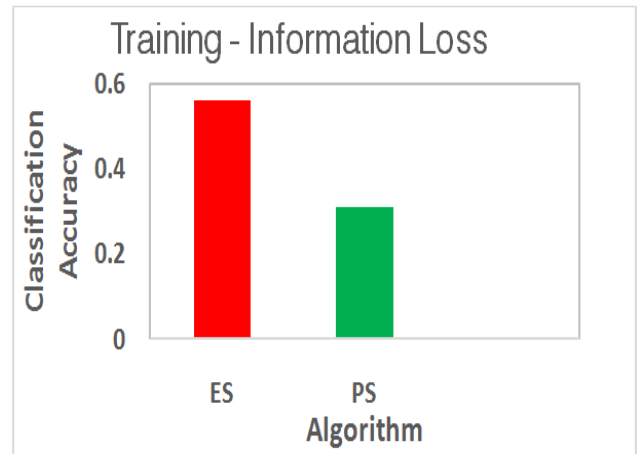


Figure 9 : Performance analysis for Information Loss for Satellite datasets

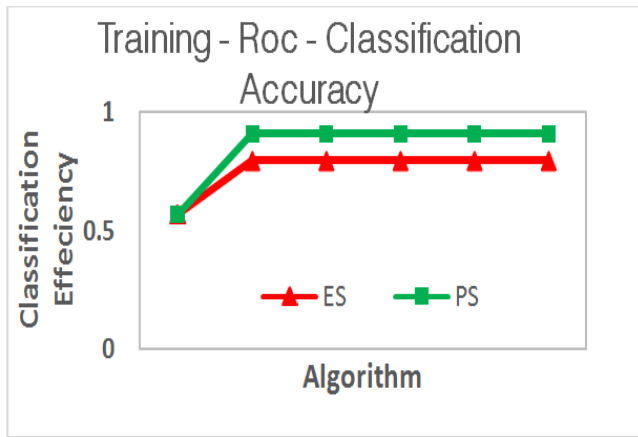


Figure 10 : ROC Analysis for Classification accuracy for Satellite datasets

The results illustrated depicts that the proposed system exhibits optimal classification with varied data size and sample space. The proposed system has exhibited better results in terms of higher classification accuracy in training, minimum information loss and optimized classification accuracy, which do satisfy every aspects of a robust privacy preserving data mining applications.

## VI. CONCLUSION

The key requirement of an effective and robust data mining system is its security or data privacy with every participating users and optimal mining efficiency. In majority of mining applications vertically partitioned data are used predominantly. In case of vertically partitioned data along with the assurance of privacy preserving in data mining, creates a situation where the rules generated are divided among parties and then certain parties remain back with fewer rules. In such circumstances, on the basis of low classification rules, the accuracy and efficiency of mining is questionable. Considering this need to generate more rules in this paper a rule regeneration scheme was proposed which not only avoids the utilization of private information allied with other parties but also enhances the classification accuracy without any computational overheads.

The developed system *dot* cumulative dot product (P<sup>2</sup>DM-RGCD) has exhibited rule regeneration with two possible rule generation functions called cumulative rule updates and dot product rule update. Using the derived functions the rule regeneration has been accomplished that makes this system highly robust to generate accurate and precise outcomes and classification accuracy. The developed system has exhibited better results in terms of its training performance, optimal classification accuracy and minimum information loss. The performance of the developed system may ensure the optimal performance with real time mining applications which needs privacy

preserving as well as optimal classification accuracy. The further evaluation and enhancement of the system can be done for Big Data applications and online web utilities.

## REFERENCES RÉFÉRENCES REFERENCIAS

1. S Kumara Swamy; Manjula S H; K R Venugopal; Iyengar S S; L M Patnaik; "A NOVEL PPDM PROTOCOL FOR DISTRIBUTED PEER TO PEER INFORMATION SOURCES"; International Journal of Computer Engineering and Technology (IJCET); Volume 4; Issue 4; July-August (2013).
2. S Kumara Swamy; Manjula S H; K R Venugopal; Iyengar S S; L M Patnaik; "Key Compromise Resilient Privacy Provisioning in Vertically Partitioned Data", In Cyber Times International Journal of Technology and Management, ISSN:2278-7518, vol. 6, no. 1, pp.18-29, March 2013.
3. S Kumara Swamy; Manjula S H; K R Venugopal; Iyengar S S; L M Patnaik; "A Data Mining Perspective in Privacy Preserving Data Mining Systems"; International Journal of Current Engineering and Technology, Vol.4, No.2 (April 2014).
4. Dehzangi, O.; Zolghadri, M.J.; Taheri, S.; Fakhrahmad, S. M., "Efficient Fuzzy Rule Generation: A New Approach Using Data Mining Principles and Rule Weighting," Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on , vol.2, no., pp.134,139, 24-27 Aug. 2007.
5. Myung-Won Kim; JoongGeun Lee; Changwoo Min, "Efficient fuzzy rule generation based on fuzzy decision tree for data mining," Fuzzy Systems Conference Proceedings, 1999. FUZZ-IEEE '99. 1999 IEEE International , vol.3, no., pp.1223,1228 vol.3, 22-25 Aug. 1999.
6. Sabu, M. K.; Raju, G., "Rule induction using Rough Set Theory — An application in agriculture," Computer, Communication and Electrical Technology (ICCCET), 2011 International Conference on , vol., no., pp.45,49, 18-19 March 2011.
7. Ji Dan; QiuJianlin; Gu Xiang; Chen Li; He Peng, "A Synthesized Data Mining Algorithm Based on Clustering and Decision Tree," Computer and Information Technology (CIT), 2010 IEEE 10th International Conference on , vol., no., pp.2722,2728, June 29 2010-July 1 2010.
8. Trincă, D.; Rajasekaran, S., "Towards a Collusion-Resistant Algebraic Multi-Party Protocol for Privacy-Preserving Association Rule Mining in Vertically Partitioned Data," Performance, Computing, and Communications Conference, 2007. IPCCC 2007. IEEE Internationa , vol., no., pp.402,409, 11-13 April 2007.

9. KumaraSwamy, S.; Manjula, S.H.; Venugopal, K.R.; Iyengar, S.S.; Patnaik, L.M., "Association rule sharing model for privacy preservation and collaborative data mining efficiency," Engineering and Computational Sciences (RAECS), 2014 Recent Advances in , vol., no., pp.1,6, 6-8 March 2014.
10. Tran, D.H.; Wee Keong Ng; Wei Zhao, "CRYPPAR: An efficient framework for privacy preserving association rule mining over vertically partitioned data," TENCON 2009 - 2009 IEEE Region 10 Conference , vol., no., pp.1,6, 23-26 Jan. 2009.
11. Modi, C.N.; Rao, U.P.; Patel, D.R., "Maintaining privacy and data quality in privacy preserving association rule mining," Computing Communication and Networking Technologies (ICCCNT), 2010 International Conference on , vol., no., pp.1,6, 29-31 July 2010.





This page is intentionally left blank



# Removal of Power Line Interference from Electrocardiograph (ECG) Using Proposed Adaptive Filter Algorithm

By Duong Trong Luong, Nguyen Duc Thuan & Dang Huy Hoang

*Hanoi University of science and Technology, India*

**Abstract-** ECG signals in measurements are contaminated by noises including power line interference. In recent years, adaptive filters with different approaches have been investigated to remove power line interference in ECG. In this paper, an adaptive filter is proposed to cancel power line interference in ECG signals. The proposed algorithm is experimented with MIT-BIH ECG signals data base. The algorithm's results are compared with the results of other adaptive filter algorithms using Least Mean Square (LMS), Normalized Least Mean Square (NLMS) by Signal to Noise (SNR). These works are performed by LabVIEW software.

**Keywords:** ECG signal processing; adaptive filters, power line interference, least mean square.

**GJCST-C Classification :** H.3.5



*Strictly as per the compliance and regulations of:*



# Removal of Power Line Interference from Electrocardiograph (ECG) Using Proposed Adaptive Filter Algorithm

Duong Trong Luong <sup>α</sup>, Nguyen Duc Thuan <sup>σ</sup> & Dang Huy Hoang <sup>ρ</sup>

**Abstract-** ECG signals in measurements are contaminated by noises including power line interference. In recent years, adaptive filters with different approaches have been investigated to remove power line interference in ECG. In this paper, an adaptive filter is proposed to cancel power line interference in ECG signals. The proposed algorithm is experimented with MIT-BIH ECG signals data base. The algorithm's results are compared with the results of other adaptive filter algorithms using Least Mean Square (LMS), Normalized Least Mean Square (NLMS) by Signal to Noise (SNR). These works are performed by LabVIEW software.

**Keywords:** ECG signal processing; adaptive filters, power line interference, least mean square.

## I. INTRODUCTION

Electrocardiography (ECG) plays an important role in monitoring and diagnosing cardiovascular diseases. The frequency bandwidth for therapy of the ECG signal is from 0.05Hz to 100Hz, and the highest peak is about 1mV [1]. During recording of ECG signals, the ECG signals get contaminated such as power supply harmonic 50Hz or power line interference (Most country including Viet Nam use 50Hz electric system) with the amplitude approaches 50% the highest peak of the ECG signal, artifacts caused by losing the direct contact between electrodes and the skin, or by EMG- this artifact's amplitude is 10% of the highest peak's amplitude of the ECG signal, or by respiratory with noise's amplitude is 15% of the highest peak of the ECG signal at 0.3 Hz frequency [2]. In the listed noises, the power line noise affects P wave and Q wave of the ECG signal. That causes errors in arrhythmia and myocardial infarction diagnosis [3]. In recent years, a few adaptive filters with different approaches are investigated to remove the power line noise 50 Hz in the ECG signal such as design of an adaptive filter with a dynamic structure for ECG signal processing [3], adaptive filtering in ECG denoising: a comparative study [4], denoising ECG signals using adaptive filter algorithm [5], denoising ECG signals with adaptive filtering algorithm & patch based method [6], investigation of adaptive filtering for noise cancellation in ECG signals [7], designing and implementation of algorithms on Matlab for Adaptive noise cancellation from ECG signal [8], performance

comparison of adaptive filter algorithms for ECG signal Enhancement [9], performance evaluation of different adaptive filters for ECG signal processing [10]. Most of the researches use Least Mean Square (LMS) and Normalize Least Mean Square algorithm (NLMS). These algorithms enable to change filter coefficients with given order; the algorithms are quite reliable and effective with small convergent time. However, in case of the power line noise's amplitude equals to 40- 50% amplitude of the highest (QRS peak) of the ECG signal, these algorithms give not effective results, and the filtered signal still contains noise. To overcome this problem, the authors propose an adaptive filter algorithm based on Fast Fourier Transform (FFT). This algorithm is experimented with ECG database such as number of record 117 and aVL lead of patient 279/s0532 from ECG database MIT-BIH [11]. The results of this algorithm are compared with that of LMS, NLMS by SNR criterion. The process and experiments are performed by LabVIEW software.

## II. METHODOLOGY

The aim of adaptive filter based on Fast Fourier Transform (FFT) is detecting the power line noise frequency, and determining threshold of this noise's magnitude.

Fourier transform:

If  $x(n)$  is a discrete signal satisfying the condition (1),

$$\sum_{n=-\infty}^{+\infty} |x(n)| < \infty \quad (1)$$

So Fourier transform equation for  $x(n)$  is given formula (2)

$$X(\omega) = \sum_{n=-\infty}^{\infty} x(n)e^{-i\omega n} \quad (2)$$

where  $i$  is the imaging part,  $i^2 = -1$

According to the material [12], B. Widrow has shown that the adaptive filter transfer function is described as (3)

$$H(z) = \frac{z^{-2} - (2\cos\omega_0)z^{-1} + 1}{(1 - 2\beta K^2)z^{-2} - \{2(1 - \beta K^2)\cos\omega_0\}z^{-1} + 1} \quad (3)$$

where  $\beta$  is the step size of the adaptive filter;  $K$  is the magnitude of power line noise;  $\omega_0$  is the angular frequency. In this study, if  $S(n)$  is ECG signal contaminated power supply harmonic 50 Hz, so  $S(n)$  is expressed in the equation (4).

Author  $\alpha \sigma \rho$  : Hanoi university of Science and Technology.  
e-mail: luong.duongtrong@hust.edu.vn

$$S(n) = X(n) + N(n) \quad (4)$$

Where  $X(n)$  is the clean ECG signal,  $N(n)$  is the power supply harmonic 50 Hz. Based on the equations (1) and (2), (4) equation can be displayed in another form (5).

$$S(\omega) = X(\omega) + N(\omega) \quad (5)$$

The magnitude-frequency spectrum of  $X(\omega)$ ,  $N(\omega)$  and  $S(\omega)$  is shown as figure 1, figure 2, and figure 3 respectively.

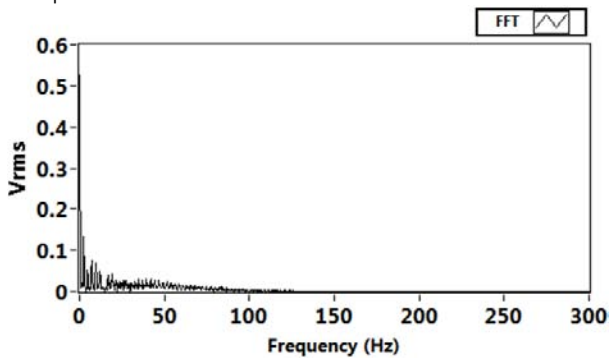


Figure 1 : Magnitude-frequency spectrum of clean ECG signal

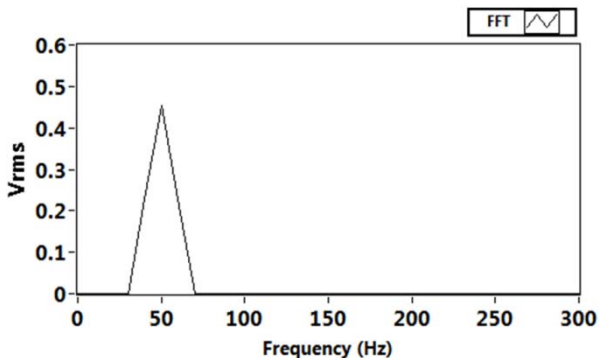


Figure 2 : Magnitude-frequency spectrum of power line noise 50 Hz with magnitude is 0.4mV

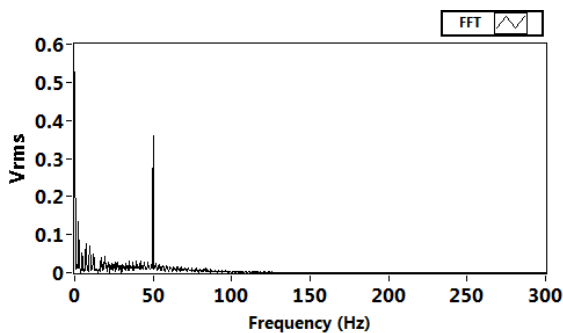


Figure 3 : Magnitude-frequency spectrum of the contaminated ECG signal

That determines the magnitude and frequency of power line noise follows these below steps.

Step 1: Importing input signal  $S(n)$

Step 2: Transforming FFT  $S(n)$  (for detecting the magnitude and frequency of the 50Hz power supply )

Step 3: Determining the frequency and magnitude of the power line interference: Choosing the frequency bandwidth is from 30Hz to 70Hz; setting the threshold of noise's FFT magnitude is 15 (based on experiments with FFT of the contaminated ECG signal)

Step 4: Initializing for loop, starting from  $i = 0$

Step 5: Checking condition: the magnitude of the power line noise in FFT is less than 15 or not? If true:

Step 6: Displaying the filtered signal and ending the processing.

If false:

Step 7: Realizing the transfer function (3)

Step 8: Calculating the output signal:

$$y(n) = H(n) \cdot S(n)$$

Step 9: Transforming FFT  $y(n)$  (for checking whether or not existing noise in the frequency bandwidth from 30Hz to 70 Hz?)

Step 10: Increasing the iteration:  $i = i + 1$  and returning to the step 5.

To experiment and test these above steps and LMS, NLMS adaptive filter algorithms, the authors used adaptive filter toolkit available in Lab VIEW. The results of the proposed algorithm are compared with that of LMS and NLMS algorithms by Signal to Noise (SNR). This criterion is followed by equation (6).

$$SNR = 20 \log_{10} \left\{ \frac{RMS(y(n))}{RMS(x(n) - y(n))} \right\} \quad (6)$$

where  $RMS(y(n))$  is the Root Mean Square of the filtered ECG signal;  $RMS(x(n))$  is the Root Mean Square of the original ECG signal.

### III. RESULTS AND DISCUSSION

The authors have tested the proposed algorithm by using a few standard ECG database records such as record mitdb117 and a VL lead of record patient 279/s 0532 from ECG database MIT-BIH added with the power supply harmonic 50 Hz having variable magnitude is from 0.4mV to 0.5mV. This signal is generated by using LabVIEW. Figure 4a is the ECG signal of the record mitdb117, and this signal is added with the power line noise 50Hz (shown in Fig.4b). Figure 4c, 4d and 4e display the results of filtering the power line interference with 0.4 mV magnitude using the proposed adaptive filter, LMS and NLMS corresponding adaptive filters. Intuitively, the filtered ECG signal using the proposed algorithm has nearly no appearance of noise and has the morphology similar to the original ECG signal.

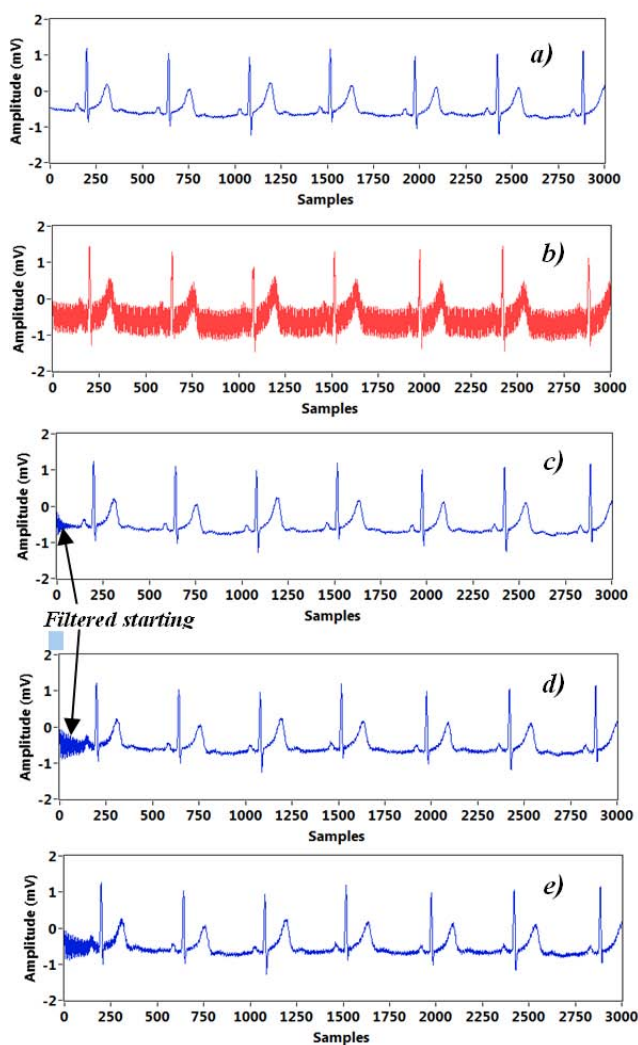


Figure 4 : The results of removing power line interference 50 Hz. a)Clean ECG signal; b) ECG signal contaminated by power line noise with 0.4 mV magnitude; c) recovered ECG signal using proposed Algorithm; d)recovered ECG signal using MLS Algorithm; e) recovered ECG signal using NLMS Algorithm

To prove the efficiency of the proposed algorithm in filtering the power line noise in ECG signal compared with LMS, and NLMS algorithms, the authors continue experimenting these algorithms with ECG data base of aVL lead of recordpatient279/s0532.The results are shown in figures 5c, 5d and 5e.These results indicate that the better efficiency of the proposed adaptive filter algorithm compared with the others. Figure 5c shows the effectiveness.

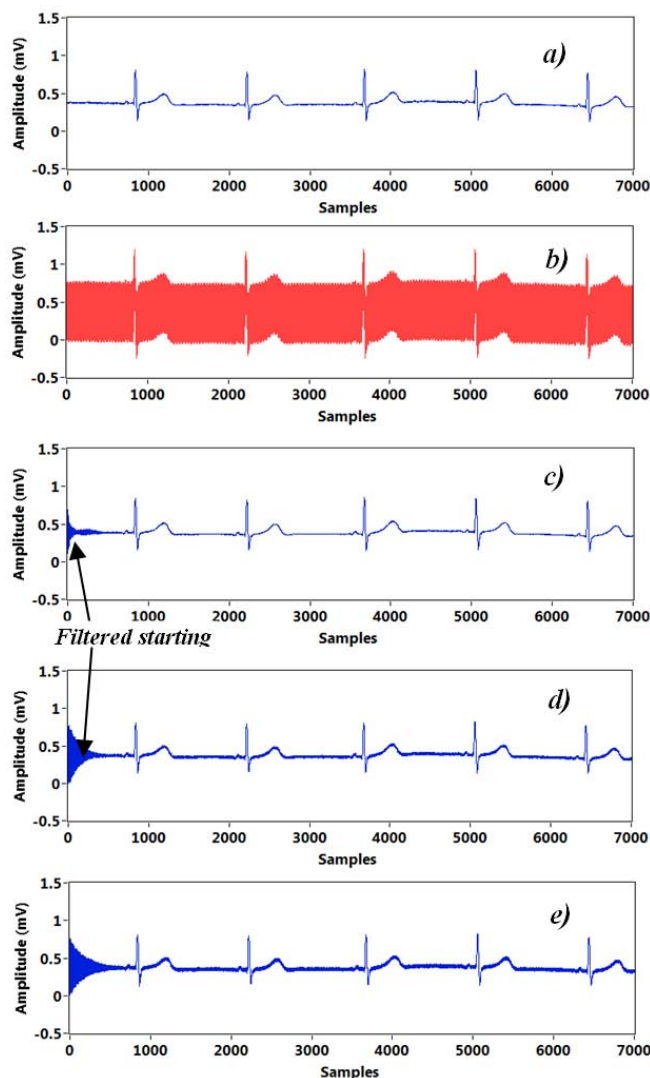


Figure 5 : The results of filtering power supply harmonic 50 Hz. a)Original ECG signal; b) Contaminated ECG signal(noise's magnitude is 0.4 mV); c) recovered ECG signal using proposed Algorithm; d)recovered ECG signal using MLS Algorithm; e) recovered ECG signal using NLMS Algorithm

To affirm the efficiency of the proposed algorithm with LMS, NLMS algorithms, the authors used SNR criterion SNR calculated by equation (6). Table 1 show the comparison among three methods filtering the power line noise with 0.4 mV magnitude, and the noise is added directly to the ECG signal in the record mitdb117.The SNR value of the proposed algorithm is higher than that of LMS and NLMS algorithms when the step size  $\beta=0.02$ ;  $\beta= 0.03$  and  $\beta=0.04$ . However, in case that  $\beta=0.01$  the SNR value of the proposed algorithm is just smaller 0.079 than that of NLMS.



**Table 1 :** Results of comparison among three algorithms filtering the power line noise 50 Hz with 0.4 mV magnitude which is added to the ECG signal of the record mitdb117.

Algorithm	Step size	SNR
LMS	0.01	6.590
NLMS		6.694
Proposed		<b>6.615</b>
LMS	0.02	6.509
NLMS		6.511
Proposed		<b>6.615</b>
LMS	0.03	5.062
NLMS		4.800
Proposed		<b>5.646</b>
LMS	0.04	5.151
NLMS		4.455
Proposed		<b>5.646</b>

**Table 2 :** Results of comparison among three algorithms which filter the power line noise (with 0.4 mV magnitude) added to ECG signal of aVL lead of record patient279/s0532.

Algorithm	Step size	SNR
LMS	0.01	2.709
NLMS		2.744
Proposed		<b>2.877</b>
LMS	0.02	2.697
NLMS		2.671
Proposed		<b>2.877</b>
LMS	0.03	2.737
NLMS		2.678
Proposed		<b>2.877</b>
LMS	0.04	2.788
NLMS		2.704
Proposed		<b>2.877</b>

From the table 2, the SNR value of the proposed algorithm is higher than that of LMS, NLMS algorithms with step size  $\beta=0.01 \div 0.04$ .

#### IV. CONCLUSION

The proposed adaptive filter algorithm for removing the power line interference in ECG signal based on Fast Fourier Transform (FFT) has been investigated with applications in the steps of the algorithm. That detects the power line frequency, and sets the threshold for the noise magnitude in FFT has high efficiency. In addition, that uses many for loops in the proposed algorithm support to filter the power line noise more carefully. The appearance of the noise is insignificant in the filtered ECG signal. Three algorithms are proceeded at the same time: adaptive filter LMS, NLMS (with different step sizes), and the proposed algorithm to remove the power supply harmonic 50 Hz (with 0.4 mV magnitude) added to the ECG signal of the record mitdb117, aVL lead of the record patient 279/s0532. The experimental results show that the proposed adaptive filter algorithm produces good ECG

signal with little noise, similar to the original ECG signal displayed in figure 4 and figure 5. Furthermore, to demonstrate the efficiency of the algorithm, the authors have compared the proposed algorithm with LMS and NLMS adaptive filter algorithms by SNR criterion. From the results in table 1 and table 2, the proposed adaptive filter algorithm for removing the power line noise 50 Hz with 0.4 mV magnitude has higher efficiency. That is asserted by SNR value in the table 1 and table 2.

#### REFERENCES RÉFÉRENCES REFERENCIAS

1. *Rangaraj M.Rangayyan* "Biomedical signal analysis". John Wiley & Sons, Inc. Canada, 2002.
2. *G.D. Clifford, F. Azuaje, and P. McSharry* "Advanced Methods and Tools for ECG Data Analysis". Norwood, MA, USA: Artech House, Inc., 2006.
3. *Ju-Won Lee, Gun-Ki Lee* " Design of an adaptive filter with a dynamic structure for ECG signal processing". International Journal of control, automation and system, vol 3, no.1, pp.137-142, 2005.
4. *I Romero, D Geng, T Berset* "adaptive filtering in ECG denoising: a comparative study". Computing in Cardiology, 39:45-48, 2012.
5. *Chinmay Chandrakar, M.K. Kowar* "Denoising ECG signals using adaptive filter algorithm". International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-1, March 2012
6. *A.Deo, M.K.Bandil, DBV Singh, A.K.Wadhvani* "denoising ECG signals with adaptive filtering algorithm & patch based method". International Journal of Computer Networks and Wireless Communications (IJCNCW), ISSN: 2250-3501, Vol.3, No3, 2013
7. *Soroor Behbahani* "Investigation of Adaptive Filtering for Noise Cancellation in ECG signals". IEEE, 2007.
8. *H.K.Gupta, R.Vijay, N.Gupta* "designing and implementation of algorithms on Matlab for Adaptive noise cancellation from ECG signal". International Journal of Computer Applications, Volume 71– No.5, 2013
9. *S.A.Rehman, R.R.Kumar* "performance comparison of adaptive filter algorithms for ECG signal Enhancement". International Journal of Advanced Research in Computer and Communication Engineering, Vol. 1, Issue 2, 2012
10. *Sachin Singh, K.L.Yadav* "performance evaluation of different adaptive filters for ECG signal processing". International Journal on Computer Science and Engineering, Vol. 02, No. 05, 1880-1883, 2010.
11. <http://physionet.org/cgi-bin/atm/ATM>
12. *B.Widrow, S.D.Steams,* "Adaptive signal processing". Englewood Cliffs, Prentice Hall. 1985



GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY: C  
SOFTWARE & DATA ENGINEERING

Volume 15 Issue 2 Version 1.0 Year 2015

Type: Double Blind Peer Reviewed International Research Journal

Publisher: Global Journals Inc. (USA)

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

# Nomenclature and Contemporary Affirmation of the Unsupervised Learning in Text and Document Mining

By Annaluri Sreenivasa Rao & Prof. S. Ramakrishna

*MREC Hyderabad, India*

**Abstract-** Document clustering is primarily a method applied for an uncomplicated, document search, analysis and review of content or is a process of automatic classification of documents of similar type categorized to relevant clusters, in a clustering hierarchy. In this paper a review of the related work in the field of document clustering from the simple techniques of word and phrase to the present complex techniques of statistical analysis, machine learning etc are illustrated with their implications for future research work.

**Keywords:** *document classification, document clustering, similarity measure, accuracy, classifiers, clustering algorithms.*

**GJCST-C Classification :** *H.2.8*



NOMENCLATURE AND CONTEMPORARY AFFIRMATION OF THE UNSUPERVISED LEARNING IN TEXT AND DOCUMENT MINING

*Strictly as per the compliance and regulations of:*



RESEARCH | DIVERSITY | ETHICS

© 2015. Annaluri Sreenivasa Rao & Prof. S. Ramakrishna. This is a research/review paper, distributed under the terms of the Creative Commons Attribution-Noncommercial 3.0 Unported License (<http://creativecommons.org/licenses/by-nc/3.0/>), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

# Nomenclature and Contemporary Affirmation of the Unsupervised Learning in Text and Document Mining

Annaluri Sreenivasa Rao<sup>α</sup> & Prof. S. Ramakrishna<sup>σ</sup>

**Abstract-** Document clustering is primarily a method applied for an uncomplicated, document search, analysis and review of content or is a process of automatic classification of documents of similar type categorized to relevant clusters, in a clustering hierarchy. In this paper a review of the related work in the field of document clustering from the simple techniques of word and phrase to the present complex techniques of statistical analysis, machine learning etc are illustrated with their implications for future research work.

**Keywords:** document classification, document clustering, similarity measure, accuracy, classifiers, clustering algorithms.

## I. INTRODUCTION

Document clustering [1], [2], [3], [4] techniques find relevance in a wide range of tasks from a simple search with a few terms to vast information retrieval processes. The early document clustering techniques used were developed for typically enhancing information retrieval systems [5], were designed to find documents according to the query type, however could not perform the task of creating a query, generate a synopsis of the documents, or provide an interface to the search results. The progress of internet, digital libraries, news sources and company-wide intranets has made available huge volumes of text documents. The tremendous increase in the already quantum size of web data and the classification of the web documents into relevant and moderate number of clusters has led to the development of large number of web clustering engines and high performing clustering algorithms.

The process of document clustering involves four stages which are,

- i) Data collection, crawling to accumulate the documents, indexing the set of documents in a structured fashion, filtering of data with techniques of tokenization, stop words removal and stemming, lemming etc.
- ii) preprocessing where the data is represented in suitable form, vector etc. and measurable factors applied to determine the similarity,

**Author α :** Associate Professor at Department of Computer Science, MRCE, Hyderabad, Telangana State, India.  
e-mail: annaluri.rao@gmail.com

**Author σ :** Prof. S. Ramakrishna Professor, Department of Computer Science, Sri Venkateswara University, Tirupathi, Andhra Pradesh, India.  
e-mail: drsramakrishna@gmail.com

- iii) Document clustering where a clustering technique and an efficient clustering algorithm are identified for clustering based on preset criteria and
- iv) Post processing involving applications of business and scientific requirements adaptation of the document clustering technique.

The applications of document clustering are of diverse nature such as,

- i) Creation of document taxonomies
- ii) IR process of search, accessing and collection [6], Similar documents identification, review and classification of results [7], automatic topic extraction [8], content summarization
- iii) Recommendation System,
- iv) Search Optimization, etc. For instance the processes are used enormously in the data classification process such as Google Web Directory, Social media data classification etc.

The clustering techniques though being studied since several years, still face many of the same challenges. These challenges [9,10] of document clustering are mostly of,

- i) Huge volume of data,
- ii) The high dimensionality of the feature space,
- iii) A feasible clustering method in terms of constraints such as cluster quality and performance and
- iv) Representing the results in an effective browsing interface. The current challenges associated with text clustering are the requirement of dynamic clustering techniques to incrementally update clusters as new data is added [11,12]. For instance the social media has to generate user specific content [13] instantly and this requires real time data clustering methodologies.

The remainder of this paper is organized as follows. In Section 2 we discuss the "Taxonomy" of document clustering, in Section 3 the "Contemporary literature work of clustering techniques" are evaluated and Section 4 gives the "Conclusion" of the paper.

## II. TAXONOMY

The clustering functionality can be expressed as a function comprising of a document set mapped to a

set of clusters. Based on specified constraints the minimum and maximum of the function defines the clustering difficulty and algorithms applied over the similarity criteria determine the clustering quality.

The preprocessing step of clustering for finding the document similarity is determined with methods based on the following strategies, (i) phrase or pair-wise methodology, (ii) tree form data depiction, (iii) component dependent data depiction, (iv) semantic relation dependent documents depiction, (v) concept and feature vector dependent depiction.

The clustering methods are generally of two types, 1) Word patterns and phrases based 2) Feature based.

The clustering methods algorithms are mostly of two types 1) hierarchical methods and 2) partitioning methods (non hierarchical) [14, 15, 16]. The hierarchical algorithms for clustering represent data sets as a cluster tree and are of two types 1-1) agglomerative [17] 1-2) divisive hierarchical clustering methods. Partitional clustering algorithms [17] are of two types, 2-1) iterative 2-2) single pass methods. K means and its variants etc. are the popular partitioning methods. The hierarchical clustering algorithms are considered efficient than the remaining algorithms [18] however due to their inherent complexity they are not applicable to huge document sets.

The techniques for determining inter-cluster similarity in classification [19 20] ex. single link and for enhancing the value of the clusters where the cluster size differs or fluctuates by a huge factor [17], especially in case of high performing clustering algorithms have been studied widely in recent years.

The widely used document clustering methods are Spectral Clustering, LSI dependent cluster development and NMF technique based clustering. The Spectral clustering methods [21] are LPI, LSI etc. Latent semantic indexing (LSI) [22] a feature extraction approach [23] tries to optimize the documents space compared to the given document and is a widely used linear document indexing method [24]. LSI is inapplicable for processes with a high range of documents [24] and similarly spectral clustering when used in a large dimensional space the dimensionality reduction is very costly which limits its usability.

The word patterns and phrases based approaches are the traditional strategies where the clustering is dependent on the documents features such as words, phrases and sequences [25, 26]. These methods are of four types, 1-1) Clustering with Frequent Word Patterns 1-2) Application of Word Clusters in Document Clusters 1-3) Co-clustering Words and Documents, Co-clustering with graph partitioning and Information- Theoretic Co-clustering 1-4) Clustering based on Frequent Phrases. The technique VSM is used in almost all the document clustering methods used nowadays [27]. The vector space model is a data model

for representing the terms related to the words in a document as a feature vector.

The features based clustering approaches are of two types 2-1) Feature Extraction 2-2) Feature Selection.

The Feature Extraction approaches are based on the algorithm of two types i) linear and ii) nonlinear techniques. The models of linear type algorithms are unsupervised PCA, OCA, MMC etc. The examples of non linear algorithms are LLE, Laplacian Eigenmaps, and ISOMAP etc. The linear methods show better operational performance in contrast to nonlinear approaches, however underperform in the clustering of huge and complicated data of the internet. The feature extraction technique finds applications in the fields of IR based on human language learning ability, comparing reviewed and submitted papers, of various languages or networks and filter of data. Feature selection algorithms are of two types, 2-2-1) Feature Ranking that is metric based and 2-2-2) Subset Selection from the possible features. The feature selection algorithms are of two categories, i) supervised and ii) unsupervised. The supervised feature selection algorithms are the most researched as well as used and they are IG, CHI, and MI. The unsupervised methods that are most popular are, i) DF-based selection dependent on term strength and ranking dependent on entropy or term contribution, ii) LSI-based method and iii) NMF based method. These techniques of unsupervised approach such as, decision trees, statistics, NLP and ML are being used in BI or analytics, in neural networks for developing AI or bio neural networks, for developing systems of AI that are rule based for intelligent content development, database development, information retrieval and automatic grouping of web documents with Enterprise Search engines or open source software's in web mining or text mining.

The strategies of feature selection used mostly are i) wrapper, ii) filter and iii) embedded methods [28] however a study [29] has shown, the methods of supervised feature selection dependent on algorithms using the filter metric IG, are most efficient over others techniques.

### III. CONTEMPORARY AFFIRMATION OF THE RECENT LITERATURE

An approach of bisecting k-means algorithm proposed by Steinbach, M, Karypis, G, & Kumar, V [14] breaks up a large cluster into small clusters repetitively to generate k numbers of clusters of huge similarity for filtering the clusters and collecting similar texts based on the method.

A technique called CCA [30] widely used in the emerging technologies of ML etc applies correlation for measuring the similar features in a document. However, CCA has its own limitations in clustering.

An approach of spectral clustering based on graph partitioning strategy called LPI [31] proposed however fails in feature selection and comprises of the existing problems of distance based clustering documents.

An approach for document clustering called Frequent Term based Clustering or HFTC [32] is a topic of extensive research. However it is not scalable for huge data or of documents.

A technique known as Hierarchical Document Clustering using Frequent itemsets (FIHC) approach proposed by Fung, B., Wang, K., Ester, M, is discussed in [33]. The strategy of FIHC though performs better than HFTC underperforms in clustering efficiency when compared to existing approaches such as UPGMA and Bisecting K-means.

The TDC algorithm technique based on closed frequent itemsets for clustering is proposed by Yu, H., Searsmith, D., Li, X., Han, J [34]. The algorithm performs better compared to HFTC and FIHC however the use of closed itemsets makes it avoidable.

A strategy of Hierarchical Clustering using Closed Interesting Itemsets, referred to as HCCI proposed by Malik, H.H., Kender, J.R [35], is the best clustering method available. However the technique may cause information loss.

An approach based on PSSM histogram by Gad and Kamel [36] combines the text semantic with the process of incremental clustering and measures the similarity of the documents for adjusting the insertion order of the documents in the cluster for quality.

An improved incremental clustering technique for an efficient clustering algorithm proposed by Gavin and Yue [37] improves categorization of web data incrementally. The method based on cluster specific multiple information anew document is assigned to a cluster.

An approach for improving text clustering mining by Shehata, S, Fakhri, K, & Mohamed S, S. [38] outperforms the existing techniques such as HAC, k-NN etc.

A progressive clustering algorithm by Liu, Y, Ouyang, Y, Sheng, H, & Xiong, Z. (2008) [39] based on Cluster Average Similarity Area determines the cluster coherence and progressively assigns the new data items to the clusters.

A technique for enhancing the clustering functionality based on the partial disambiguation of words by means of their PoS [40] is recommended by the developers as the approach finds the inefficiency of considering synonyms and hypermy my for selecting the right sense of the word disambiguated solely by PoS tags.

The CFWS technique proposed by Y. LI, and S.M. Chung, enhances the capability to process the document, considering the word sequences apart from the words [41].

The technique of non linear representation of the data by J.B. Tenenbaum, V. de Silva, and J.C. Langford [42] keeps specific local data simultaneously based on the optimization factors however is associated with high complexity.

A study of the approaches for reducing the complexity of feature extraction based on a new technique called approximation algorithm [43], [44], [45] is found to be good.

A software for automatically retrieving information from websites by Zamir O Etzioni [46] is designed for websites comprising of vast amount of data

The approach of integrating clustering and feature selection for text clustering based on the semantic relation of the text documents with ontology was proposed by Thangamani.M and P.Thangaraj in [47]. The approach minimizes dimensionality and improves feature selection.

The clustering technique, for finding the clustering quality based on WordNet [48] phrasal noun and semantic relationships [49] shows better performance with hyperny my based strategy compared to other noun phrases.

A system for determining the ontology related semantic relations of the term or word and associated weight measure is given by Prof. K. Raja, C. Prakash Narayanan [9]. However the technique has dimensionality and other problems.

A description of the task of Ontology based automatic categorizing of web documents [50] and the scope of Ontology in improving the current machine learning and IR approaches is given by Andreas Hotho. The integration of ontology's for combining various information types of multiple resources by Young-Woo et al. in the paper [51].

The process of using domain specific ontology's for enhancing performance of text classification where text learning and IR are used to generate ontology's with minimum user interaction is given in [52, 53].

The methods utilizing Wikipedia ontology for improving primarily the document depiction and cluster quality by Gabrilovich and Markovitch [54] and a further extension provided a structure based on the Wikipedia guidelines and groups [55, 56]. The Wikipedia ontology is most relevant as it is applicable to a large cross section of domains and also restructured on a regular basis.

A technique for feature selection in text clustering based on supervised feature selection on the intermediary clustering outcomes by Xu, J. Xu, B [57] generates a efficient subset for classification. The suggested techniques performance is efficient compared to manual process.

A technique of feature selection dependent on the ACO algorithm by M. Janaki Meena,K.R.

Chandran, J. Mary Brinda," [58] is a unique method. Comparative tests of the approach with existing chi-square and CHIR techniques shows the proposed approach achieves better performance in FS.

An entropy based FS approach i.e. a filter solution [59] tested with various data types that reduces dimensionality and is efficient in finding the subset of major features.

A feature co-selection method called MFCC (multi type feature co-selection), proposed by Shen huang, Zheng Chen, Yong Yu, and Wei-Ying main [60] shows enhanced clusters performance of web documents based on the outcomes of intermediate clustering.

A method to remodel the matrix of data similarity as a bi-stochastic matrix prior to executing algorithms by F. Wang, P. Li, and A. C. K Aonig showed better clustering performance [61].

The techniques of document clustering that are term based for clustering in dynamic environments, is given in [11] by Wang, X, Tang, J, & Liu, H, synonyms and hypermy my by Bharathi and Vengatesan [62], Synonyms and Hyponyms, Nadig, R, Ramanand, J, & Bhattacharyya, P in [12]. These approaches are however not applicable to technically similar documents.

A document clustering approach [63] dependent on phrases and the STC technique by O. Zamir, O. Etzioni, O. Madanim, and R.M. Karp builds the clusters on the common documents suffixes. The method though efficient in cluster quality however is associated with high amount of term redundancy.

A study of the TF-IDF method of clustering [64], term frequency dependent algorithms [65] and a review of clustering algorithms [66] showed that majority of clustering approaches are TF-IDF based, however associated with several problems.

The NMF (Nonnegative Matrix Factorization) technique in text classification [67], improved clustering performance compared to the existing approaches [68], relationship study of NMF techniques with earlier clustering techniques [69], [70] [71]. A review of established techniques of NMF such as multiplicative updates [72], projected gradients [73] though efficient however are associated with the problems of memory for huge datasets streamed and not disk based [74]. To overcome these problems, approaches such as random projections [61, 75] and sketch/sampling algorithms [76] have been proposed. An NMF based technique by Li and Zhu in 2011 [77] for research specific documents minimizes high dimensionality, finds relevant topics for clustering and shows performance efficiency in classification comparatively. A study of the online algorithm based on Nonnegative Matrix Factorization [78], a NMF based method that uses features based on weights and similar cluster property by Sun Park, Dong Un An, Choi Im Cheon [79] performs comparatively

more efficiently than the remaining NMF based strategies.

#### IV. CONCLUSION

In this paper we analyzed several techniques developed for clustering documents with their applications and relevance in terms of today's requirements. The task of developing perfect strategies for classification of varied forms and types of documents for a near optimal solution or finding accurate ways of assessing the quality of the performed clustering though is impossible and is increasing in its complex nature, the field today deals with extraordinary tasks like granular taxonomies generation, sentiment analysis and document summarization for generating reliable and relevant insights applicable to several fields. In conclusion we can say document clustering is going to be widely studied and will find relevance in a number of newer areas.

#### REFERENCES RÉFÉRENCES REFERENCIAS

1. S. Kotsiantis and P. Pintelas, "Recent Advances in Clustering: A Brief Survey," WSEAS Trans. Information Science and Applications, vol. 1, no. 1, pp. 73-81, 2004.
2. W. Xu and Y. Gong, "Document Clustering by Concept Factorization," Proc. Int'l Conf. Research and Development in Information Retrieval, pp. 202-209, July 2004.
3. S. Siersdorfer and S. Sizov, "Restrictive Clustering and Metaclustering for Self-Organizing Document Collections," Proc. Int'l Conf. Research and Development in Information Retrieval, pp. 226-233, July 2004.
4. Brian S. Everitt, Sabine Landau, and Morven Leese. Cluster Analysis. Oxford University Press, fourth edition, 2001.
5. Van Rijsbergen, C.J. Information Retrieval (Secondth ed.). London: Buttersworth. 1989.
6. C. Carpineto, S. Osiński, G. Romano, and D. Weiss. A survey of web clustering engines. ACM Comput. Surv., 41(3):1-38, 2009.
7. X. Liu and W. B. Croft. Cluster-based retrieval using language models. In Proceedings of the 27th annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pages 186-193, 2004.
8. J. Silva, J. Mexia, A. Coelho, and G. Lopes. Document clustering and cluster topic extraction in multilingual corpora. In Proceedings of the 1st IEEE International Conference on Data Mining (ICDM), pages 513-520, 2001.
9. Prof. K. Raja, C. Prakash Narayanan, "Clustering Technique with Feature Selection for Text Documents", Proceedings of the Int.Conf. on

- Information Science and Applications ICISA 2010 6 February 2010, Chennai, India.
10. Fabrizio Sebastiani "Machine Learning in Automated Text Categorization" ACM Computing Surveys, Vol. 34, No. 1, March 2002
  11. Wang, X, Tang, J,& Liu, H Document clustering via matrix representation. In 11th IEEE International Conference on DataMiningICDM2011 (pp. 804–813), 2011.
  12. Nadig, R, Ramanand, J, & Bhattacharyya, P. (2008).Automatic evaluation of Word Net synonyms and hypermy my India: Proceedings of ICON-2008, 6th International Conference on Natural Language Processing.
  13. A. Das, M. Datar, A. Garg, and S. Rajaram. Google news personalization: Scalable online collaborative filtering. In Proceedings of the 16th International Conference on World Wide Web (WWW), pages 271{280, 2007.
  14. M. Steinbach, G. Karypis, and V. Kumar. A comparison of document clustering techniques. KDD Workshop on Text Mining" 2000.
  15. P. Berkhin. 2004. Survey of clustering data mining techniques [Online]. Available: [http://www.accrue.com/products/rp\\_cluster\\_review.pdf](http://www.accrue.com/products/rp_cluster_review.pdf).
  16. Xu Rui.. Survey of Clustering Algorithms. IEEE Transactions on Neural Networks, 16(3):pp. 634-678, 2005.
  17. B.C.M.Fung, K.Wan, M.Ester. 2003. Hierarchical Document Clustering Using Frequent Itemsets", SDM"03.
  18. I. S. Dhillon and D. S. Modha. Concept decompositions for large sparse text data using clustering. Machine Learning, 42(1):143{175, 2001.
  19. M. Khalilian & N. Mustapha (2010), "Data Stream Clustering: Challenges and Issues", In Proceedings of the International Multiconference of Engineers and Computer Scientists IMECS 2010, Hong Kong, pp. 978-988.
  20. R. Martinez-Morais, F.J. Alfaro-Cortes, & J. L. Sanchez (2010), "Providing QoS with the Deficit Table Scheduler", IEEE Transactions on Parallel and Distributed Systems, Vol. 21, No. 3, pp. 327-341.
  21. A.Y. Ng, M. Jordan, and Y. Weiss, "On Spectral Clustering: Analysis and an Algorithm," Advances in Neural Information Processing Systems 14,pp. 849-856, Cambridge, Mass.: MIT Press,2001.
  22. S.T. Dumais, "Latent Semantic Indexing (LSI) and TREC-2,"Proc.Second Text Retrieval Conf. (TREC),pp. 105-116, 1993.
  23. LSA @ CU Boulder, <http://lsa.colorado.edu/>, 2010.
  24. S.C. Deerwester, S.T. Dumais, T.K. Landauer, G.W. Furnas, and R.A. Harshman, "Indexing by Latent Semantic Analysis,"J. Am.Soc. Information Science, vol. 41, no. 6, pp. 391-407, 1990.
  25. Hung, C. and Xiaotie, D., "Efficient Phrase-Based Document Similarity for Clustering," IEEE Transaction on Knowledge and Data Engineering, vol. 20, no. September, pp. 1217- 1229, 2008.
  26. Soon, M. C. , John, D. H., and Yanjun, L., "Text document clustering based on frequent word meaning sequences," Data & Knowledge Engineering, vol. 64, pp. 381-404, 2008.
  27. Aas, K,& Eikvil, L.(1999). Text Categorisation: A Survey. Technical Report 941. Oslo Norway: Norwegian Computing Center.iteseer.ist.psu.edu/aas99text.html.
  28. Barak Chizi (Tel-Aviv University, Israel); Lior Rokach (Ben-Gurion University, Israel); Oded Maimon (Tel-Aviv University, Israel) "a survey of feature selection techniques"1888-1895 pp. John Wang (Ed.) (Montclair State University, USA), DOI:10.4018/978-1-60566-010-3.ch289, ISBN13:9781605660103, 2009.
  29. Y. Yang and J.O. Pedersen, "A Comparative Study on Feature Selection in Text Categorization," Proc. 14th Int'l Conf. Machine Learning, pp. 412-420, 1997.
  30. D.R. Hardoon, S.R. Szedmak, and J.R. Shawetaylor, "Canonical Correlation Analysis: An Overview with Application to Learning Methods,"J. Neural Computation, vol. 16, no. 12, pp. 2639-2664, 2004.
  31. D. Cai, X. He, and J. Han, "Document Clustering Using Knowledge and Data Eng.,vol. Locality Preserving Indexing,"IEEE Trans. 17, no. 12, pp. 1624-1637, Dec. 2005.
  32. Beil, F., Ester, M., Xu, X.: Frequent Term-based Text Clustering, In Proc. of Intl. Conf. on Knowledge Discovery and Data Mining,. (2002).
  33. Fung, B.C.M., Wang, K., and Ester, M., "Hierarchical document clustering using frequent Itemsets," Proceedings of SIAM International Conference on Data Mining, 2003.
  34. Yu, H., Searsmith, D., Li, X., Han, J.: Scalable Construction of Topic Directory with Nonparametric Closed Termset Mining, In Proc. of Fourth IEEE Intl. Conf.on Data Mining. (2004).
  35. Malik, H.H., Kender, J.R.: High Quality, Efficient Hierarchical Document Clustering Using Closed Interesting Itemsets, In Proc. of IEEE Intl. Conf. on Data Mining. (2006).
  36. Gad, WK, & Kamel, MS. (2010). Incremental clustering algorithm based on phrase- semantic similarity histogram. Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, 11(14), 2088–2093.
  37. Gavin, S, & Yue, X. (2009). Enhancing an incremental clustering algorithm for Web page collections. IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technologies,81–84.

38. Shehata, S, Fakhri, K, & Mohamed S, S. (2010). An efficient concept-based mining model for enhancing text clustering. *IEEE Transactions On Knowledge And Data Engineering*, 22(10), 1360–137.
39. iu, Y, Ouyang, Y, Sheng, H, & Xiong, Z. (2008). An Incremental Algorithm for Clustering Search Results, *IEEE International Conference on Signal Image Technology and Internet Based Systems*(pp. 112–117).
40. Sedding, J., & Kazakov, D. 2004. Wordnet-based text document clustering. 3rd Workshop on Robust Methods in Analysis of Natural Language Data, pp. 104–113.
41. Y. LI, and S.M. Chung. 2005. Text Document Clustering Based on Frequent Word Sequences. In *Proceedings of the. CIKM, 2005. Bremen, Germany, October 31- November 5*.
42. J.B. Tenenbaum, V. de Silva, and J.C. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," *Science*, vol. 290, pp. 2319-2323, 2009.
43. J. Weng, Y. Zhang, and W.-S. Hwang, "Candid Covariance-Free Incremental Principal Component Analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 1034-1040, Aug.2003.
44. K. Hiraoka and M. Hamahira, "On Successive Learning Type Algorithm for Linear Discriminant Analysis," *IEIC Technical Report*, (in Japanese), vol. 99, pp. 85-92, 1999.
45. J. Yan, Zhang, B.S. Yan, Z. Chen, W. Fan, Q. Yang, W.Y. Ma, and Q. Cheng, "IMMC: Incremental Maximum, Marginal Criterion," *Proc. 10th ACM SIGKDD*, pp. 725-730, 2004.
46. Zamir, O. Etzioni, "Web Document Clustering, A Feasibility Demonstration," in *Proceedings of the 21st International ACM SIGIR Conference on Research and Development*. K. Mugunthadevi et al. / *International Journal on Computer Science and Engineering (IJCSSE)* ISSN :
47. Thangamani.M and P.Thangaraj," integrated clustering and feature selection scheme fo textdocuments", *J.Comput.Sci.*,6:536.541, DOL:10.3844/jcssp.2010.536.541, URL:<http://www.thescipub.com/abstract/10.3844/jcssp.2010.536.54>.
48. Miller G. 1995. Wordnet: A lexical database for English. *CACM*, 38(11), pp. 39–41.
49. Zheng, Kang, Kim. 2009. Exploiting noun phrases and semantic relationships for text document clustering. *Information Science* 179 pp. 2249-2262.
50. Andreas Hotho," Using Ontologies to Improve the Text Clustering and Classification Task", *Knowledge and Data Engineering Group, University of Kassel, January 14, 2005*.
51. Young-Woo Seo, Anupriya Ankolekar, Katia Sycara," Feature Selections for Extracting Semantically Rich Word for Ontology Learning" *CMU-RI-TR-04-18. March 2004*.
52. B. Berendt, A. Hotho, and G. Stumme, "Towards semantic web mining", In *Proceedings of International Semantic Web Conference (ISWC)*, pages 264– 278,2002.
53. A. Hotho, S. Staab, and A. Maedche. Ontology-based text clustering. In *Proceedings of the IJCAI-2001 Workshop Text Learning: Beyond Supervision, Seattle, USA, August 2001*.
54. Gabilovich, E., Markovitch, S.: Computing Semantic Relatedness Using Wikipedia-based Explicit Semantic Analysis, In *Proc. of The 20th Intl. Joint Conf.on Artificial Intelligence. (2007)*
55. Hu, X., Zhang, X., Lu, C., et al.: Exploiting Wikipedia as External Knowledge for Document Clustering, In *Proc. of Knowledge Discovery and Data Mining. (2009)*.
56. Hu, J., Fang, L., Cao, Y., et al: Enhancing Text Clustering by Leveraging Wikipedia Semantics, In *Proc. of 31st Annual Intl. ACM SIGIR Conf. on Research and Development in Information Retrieval. (2008)*
57. Xu, J. Xu, B. Zhang, W. Cui, Z. Zhang, W."A new feature selection method for text clustering ", *wuhan university journal of natural sciences*, 2007, vol 12; number 5, pages 912-916.
58. M. Janaki Meena, K.R. Chandran, J. Mary Brinda," integrating swarm intelligence and statistical data for feature selection in text categorization" ©2010 *International Journal of Computer Applications (0975 – 8887)*, Volume 1 – No. 11.
59. Manoranjan Dash ,Kiseok Choi ,Peter Scheuermann ,Huan Liu," Feature Selection for Clustering – A Filter Solution" *Proceedings of the 2002 IEEE International Conference on Data Mining (ICDM'02)0-7695-1754-4/02 © 2002 IEEE*.
60. shen huang, zheng chen, yong yu, and wei-ying ma," multitype features coselection for web document clustering", *ieee transactions on knowledge and data engineering*, vol. 18, no. 4, april 2006, 1041-4347/06/\$20.00 \_ 2006 *ieee published by the ieee computer society*.
61. F. Wang, P. Li, and A. C. KÄonig. Learning a bistochastic data similarity matrix. In *Proceedings of the 10th IEEE International Conference on Data Mining (ICDM)*, 2010.
62. Bharathi, G, & Vengatesan, D. (2012). Improving information retrieval using document clusters and semantic synonym extraction. *Journal of Theoretical and Applied Information Technology*, 36(2), 167–173.
63. O. Zamir and O. Etzioni, "Grouper: A Dynamic Clustering Interface to Web Search Results," *Computer Networks*, vol. 31, nos. 11-16, pp. 1361-1374, 1999.



64. Salton, G, & Buckley, C. (1998). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5), 513–523.
65. Kumar, N, & Srinathan, K. (2009). A New Approach for Clustering Variable Length Documents (Proceedings of the Advanced computing Conference, IEEE, pp. 982–987).
66. Prathima, Y, & Supreethi, KP. (2011). A survey paper on concept based text clustering. *International Journal of Research in IT & Management*, 1(3), 45–60.
67. D. D. Lee and H. S. Seung. Learning the parts of objects with nonnegative matrix factorization. *Nature*, 401: 788-791, 1999.
68. F. Shahnaz, M. W. Berry, V. P. Pauca, and R. J. Plemmons. Document clustering using nonnegative matrix factorization. *Information Processing and Management*, 42(2):373-386, 2006.
69. C. Ding, T. Li, and M. I. Jordan. Convex and seminonnegative matrix factorizations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010.
70. C. Ding, X. He, and H. D. Simon. On the equivalence of nonnegative matrix factorization and spectral clustering. In *Proceedings of the 5th SIAM Int'l Conf. Data Mining (SDM)*, pages 606- 610, 2005.
71. E. Gaussier and C. Goutte. Relation between pls and nmf and implications. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, pages 601-602, 2005.
72. D. D. Lee and H. S. Seung. Algorithms for nonnegative matrix factorization. In *Advances in Neural Information Processing System (NIPS)*, pages 556-562, 2000.
73. C. J. Lin. Projected gradient methods for nonnegative matrix factorization. *Neural Computation*, 19(10):2756 - 2779.
74. S. Zhong. Efficient streaming text clustering. *Neural Networks*, 18(5-6):790{798, 2005.
75. F. Wang and P. Li. efficient non-negative matrix factorization with random projections. In *Proceedings of the 10th SIAM International Conference on Data Mining (SDM)*, pages 281-292, 2010.
76. P. Li, K. W. Church, and T. Hastie. One sketch for all: Theory and application of conditional random sampling. In *Advances in Neural Information Processing System (NIPS)*, pages 953-960, 2008.
77. Li, F, & Zhu, Q. (2011). Document clustering in research literature based on NMF and tensor theory. *Journal of Software*, 6(1), 78–82.
78. B. Cao, D. Shen, J. Sun, X. Wang, Q. Yang, and Z. Chen. Detect and track latent factors with online nonnegative matrix factorization. In *Proc. International Joint Conference on Artificial Intelligence*, pages 2689- 2694, 2007.
79. Sun Park, Dong Un An, Choi Im Cheon, "Document Clustering Method Using Weighted Semantic Features and Cluster Similarity," *digitel*, pp.185-187, 2010 Third IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning, 2010.



This page is intentionally left blank



# Changing Scenario of Testing Paradigms Using DevOps - A Comparative Study with Classical Models

By Surendra Naidu Mullaguru

*JNTUK, India*

**Abstract-** Many organizations come across themselves in a challenging conditions with the increase of web applications which depend on cloud. As cloud applications are supportive for rapid releases with respect to the queries requested by the user community. Most of the software organizations expects to release the product in time as required by the end user hence, it pressurize functional teams of the organization to do the needful. As perfection is rare to achieve , it is obvious to commit more and more defects and consequent annoyance to the team. This paper discusses about a new paradigm i.e DevOps which tries to address the identified problem by integrating Development and Operations teams. In this model, the development team would take care of operational requirements like deploying scripts, debugging and performance testing from the scratch and the operations team take care of well-informed support and feedback before, during, and after deployment.

*GJCST-C Classification : D.2.5*



*Strictly as per the compliance and regulations of:*



# Changing Scenario of Testing Paradigms Using DevOps – A Comparative Study with Classical Models

Surendra Naidu Mullaguru

**Abstract-** Many organizations come across themselves in a challenging conditions with the increase of web applications which depend on cloud. As cloud applications are supportive for rapid releases with respect to the queries requested by the user community. Most of the software organizations expects to release the product in time as required by the end user hence, it pressurize functional teams of the organization to do the needful. As perfection is rare to achieve, it is obvious to commit more and more defects and consequent annoyance to the team. This paper discusses about a new paradigm i.e DevOps which tries to address the identified problem by integrating Development and Operations teams. In this model, the development team would take care of operational requirements like deploying scripts, debugging and performance testing from the scratch and the operations team take care of well-informed support and feedback before, during, and after deployment.

## I. INTRODUCTION

Most of the software development teams are moving towards a new trend of testing domain i.e DevOps. Organizations are given pressures to produce higher-quality code faster giving less time for Quality Assurance (QA). As it is emerging as prominent domain, the developers need to adjust themselves to gain more benefits. We cannot release the product quickly with timelines as persisting in compact and separated work teams. The DevOps promises to achieve this. With DevOps, team's participating skills are equally important as technical skills. This paper focus on how DevOps paradigm can change the testing set-up of conventional testing by integrating the development process and the QA with the operations functioning together to speed up development during problem resolution. The paper also focuses on the areas of application, its suitability wherein the conventional system fail to deliver testing inferences [7]. DevOps is a rising methodology from two popular methods of software development. i.e the agile system administration or agile operations and the later is the integration of development and operations teams in all the phases of the development lifecycle for creating and using a service. The definition of DevOps according to Jez Humble is "DevOps is a cross-disciplinary community of practice dedicated to the learn building,

evolving and operating rapidly-changing resilient systems at scale". DevOps is an effort to enable operations engineers and development engineers together in total service lifecycle, from design through the development process to production phase and support. DevOps is also characterized by operations by team making use of many of the techniques as developers to carry out systems work in a routine pattern. Those techniques can range from using source control to testing to participate in an agile development process. In the industry the tag DevOps is referred as the relationship between development and operations. The reason for this is to get the advantage of both development team and operations team together for the growth of the business. The DevOps improves the organization potential for nonstop software release that can facilitate organizations to capture market opportunities. It also helps in evaluating customer feedback, cost, quality and risk. As DevOps analyzes the code systematically and assist in reducing the coding errors. According to the present market trend in the near future DevOps is expected to emerge as a vital tool for rapid process develop. In the past development employees are considered as "makers" and the "Ops" employees are considered as "people who deal with the creation after its birth. In this way, DevOps can be implemented as a outcome of Agile Computing. Agile software development stipulates close association of customers, product management, developers, and QA to fill in the space and rapidly iterate towards an improved product. But the aim of DevOps is not only increasing the rate of change but also to deploy the features successfully into production without causing disorder and trouble to other services, while rapidly detecting and correcting incidents when they occur. DevOps is a simple extension to agile principles beyond the boundaries of "the code" to the entire delivered service. DevOps is especially complementary to the agile software development process, as it extends and completes the continuous integration and release process by ensuring the code production and provide value to the customer[6][1].

The testing tool must ensure to provide some basic provision to carry out testing facilities which are being considered by DevOps as a vital aid. Following are the some basic capabilities extend by DevOps[4].

*Author:* Sharp Software Development India Pvt.Ltd, Test Lead, India  
*e-mail:* surendranaidu0580@gmail.com

- Testing early pieces of code often.
- Integration of application development and test life cycles.
- Standardization of the testing methods for consistent results.
- Develop a broad test plan, as it creates the basis for the testing methodology.
- Extend dynamic approach and declare expected results.
- To manage business policies across the application.
- Usage of multiple levels and kinds of testing (systems testing, regression, module, integration, stress and load).
- Assessment and examining the tasks to lower costs and for plan for future.
- Easy interface for the clients to verify the extent of application development and feasibility.

## II. COMPARISON OF DEVOPS WITH AGILE

The origin for the development of DevOps is emerging technologies. The DevOps is derived from the Agile System Administration movement and Enterprise Systems Management (ESM) movement. The characteristic of agile process is that it shows us how to create adaptive software development for quick feedback cycles through early customer participation. The primary characteristic of DevOps is that it contemplates on business results rather than technical details. DevOps acquires lot of lean principles like deploying in smaller batches aiming at enhancing the communication between developers and operations teams[3][2].

### a) *Lean is the basis of Agile*

Lean signifies optimization from end-to-end process which produces value for your customer – from the original idea to generation of revenue. Lean philosophy focus on stream more than bottlenecks in the process and unproductive activities identification. Creating importance for your customers is all about judgment, validate product ideas, implementing these ideas and managing them to advantage the users. It is evident that many admirable schools of thought are there for structuring the processes such as Lean Startup and Running Lean to hash out product ideas. Scrum, and XP are used for creating features in a pliable manner. The DevOps enables the creation of a tradition of focusing on delivering worth for the customer instead of only concerning for technical issues[1][4].

### b) *Lean Startup or Running Lean*

Eric Ries and Ash Maurya reviewed about a learning cycle during feature discovery. From authenticating ideas to testing probable solutions, they

teach to resolve problems worth solving and keep your solutions more directed.

- *Scrum, XP and Kanban:* Typical agile processes show how to arrange adaptive software development for speedy feedback cycles through timely customer participation.
- *DevOps:* It focus on cultural aspects of business results relatively than technical details. DevOps integrate lot of lean principles, focusing on improvement in collaboration between developers and operations folks.

Lean principles are the foundation of entire product development course. In order to incorporate path of improvement, one must consider the whole as sub-optimizing element and just one part will hurt the business. The developer circles are emphasizing on DevOps and the entrepreneur circles are focusing at Lean Startup. The term, coined by Eric Ries, is a startup that achieves validated learning through the build-measure-learn loop. With a notion of learning what the customer actually need so that a useful product and a profitable business model would emerge[2][8].

It may not appear that the Lean Startup and DevOps are related. Small batches of tasks are the way for cross-functional teams, which work together on the smallest units of work possible while pushing that work to the customer as soon as possible. This way, the product gets to the customer swiftly, allowing the business model or product to be validated.

The common model presently is large batches. Instead of a small team working across the modules to push a product towards the customer rapidly. The team specializes, with a range of team members completing their segment of the product before passing it on to the next phase in the chain. If at any point a fault is found, the product has be sent back down to the chain. This leads to speedy feedback and rapid turnaround times, resulting in an improved product. DevOps is about accepting small batches, but the batch in this case is a thin part of the overall process, pushing a product from development to deployment to see it in use. DevOps is about allowing production behavior to quickly fed back to developers so that development can adjust swiftly to the situations in production. With the lean startup, the potential competence of an individual unit such as operations might emerge to weaken. The efficiency of the system as a whole speeds up. When developing a system it is simple to foresee how a system should run, but that is not enough. Often inevitable differences in hardware and traffic escort to differences in how the system actually performs with how it was predicted. DevOps is about tightening the same loop as the Lean Startup. The only difference is the narrower focus of DevOps. With the loop as rigid as possible, knowledge gained from the real system is fed back into the next cycle of the product so the system becomes better in

small and tight increments. DevOps is a subset of the Learn Startup. Both are reactions to the today's fast changing world and require to produce a system that can optimize quickly, even if that means undercutting what appears to be the efficient in the entity unit.

c) *Jenkins – A DevOps tool*

Jenkins is a rapid Continuous Integration server. Continuous integration is a procedure in which all development work is integrated at as planned entity with a time constraint or event and the resulting work is routinely tested and built in automated environment as shown in figure 1. The idea is to development a scenario wherein errors are being identified at an early stage in the process. Jenkins is an open source tool to achieve continuous integration. The basic functionality of Jenkins is to perform a definite list of steps supported by a trigger. The trigger might be a change in version control system or a indication for time based trigger, e.g., a build every 15 minutes. To accomplish this the list of steps include:

- Perform a software build with Apache Maven or Gradle
- Run a shell script
- Archive the build results
- Start the integration tests

Basically Continuous Integration is the mechanism of running tests on a non-developer machine routinely each time someone pushes a fresh code into the source repository.

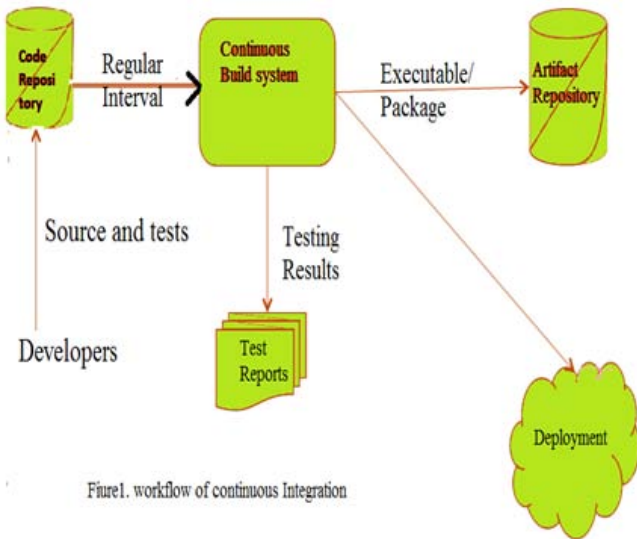


Figure 1 : Widov of continuous integration

This has remarkable advantage of knowing that all the tests work and attain fast feedback. The fast feedback is vital to know right after breaking the build and based on that one would repair or revert it for improvement. The developer must be alert when running the tests occasionally, as the problem that has emerged due to change in code may have surfaced since the

last time. As it is hard to figure out which change has induced such problem. Built on continuous Integration is the platform of continuous deployment/ delivery wherein, after a successful test run the system instantly would release the latest version of the codebase. This makes the deployment a non-issue and helps in pace up development. Jenkins is a one of the DevOps application tool which supervises regularly executed jobs, such as building a software project. Among those things, presently Jenkins focuses on building/testing software projects continuously and supervise executions of externally-run jobs, such as cron jobs and procmail jobs, even though that are run on a remote machine[5].

The following section demonstrates the steps of a sample Java application testing using Jenkins tool. The implementation steps are as follows.  
 step1: Deploy Jenkins. war in web server  
 step2: Create a new job.  
 step3: Add build steps in job configuration.  
 step4: From the job home page click on “Build Now” to start the build process.  
 step5: Check the log for build status and then click1 on last run console output.

The following are the screenshots of a simple java application testing using jenkins tool. After deplyong the Jenkins. war in web server we create a new job. Then add build steps in job configuration. Once it is configured, we would get a resultant screen as figure 2.

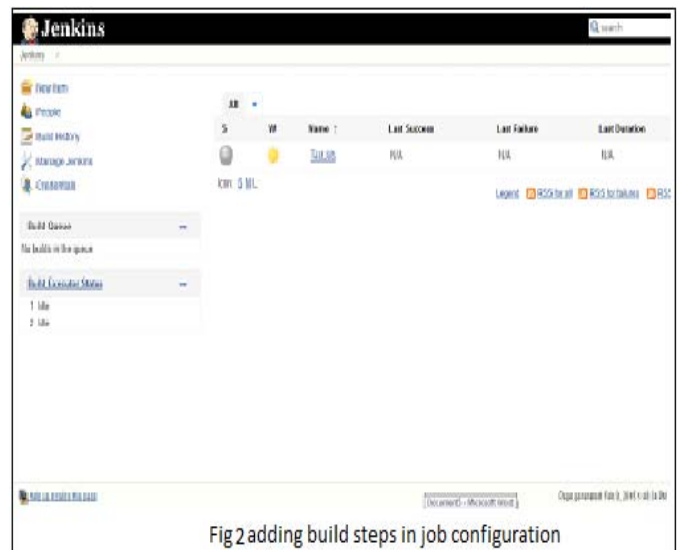


Figure 2 : Adding build steps in job configuration

Once the build steps are added in the job configuration we started the build process which is depicted in the figure 3 and the corresponding log report is depicted in figure 4.

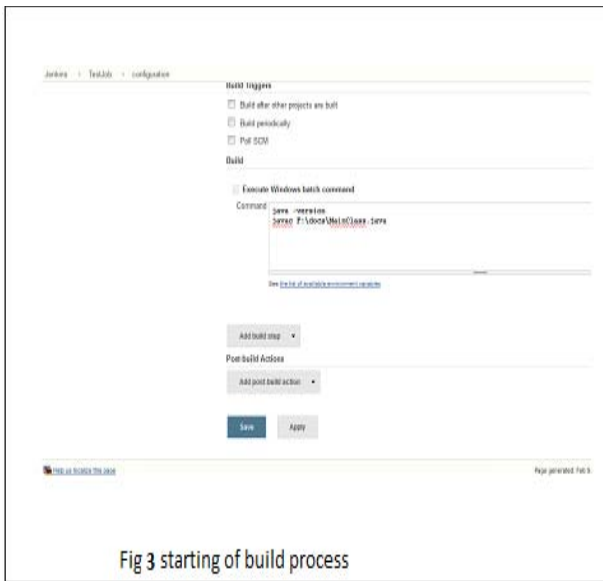


Fig 3 starting of build process

Figure 3 : Starting of build process.

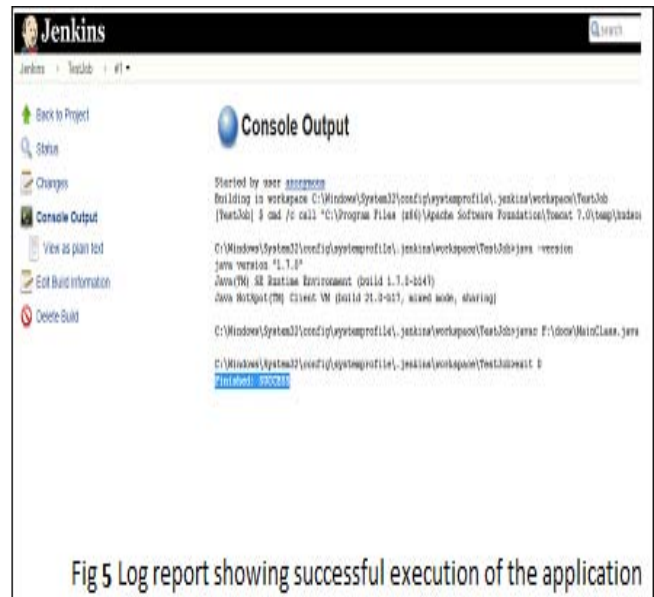


Fig 5 Log report showing successful execution of the application

Figure 5 : Log report showing successful execution of the application.

Figure 5 shows the successful execution of the application. In case if the java application is modified then automatically jenkins will trigger error report which is shown in the form of log report as depicted in the figure 6.



Fig 4 log report for build status

Figure 4 : log report for build status.



Fig 6 Log showing error report

Figure 6 : Log showing error report

### III. CONCLUSION

Automated test data generation is a central area of research for sinking cost of software development. DevOps is definitely a new approach in testing strategies which increases the organization throughput. Based on the test information requirements the test data generation methods, diverse program

analyzers are utilized To face the today's cloud based environment DevOps is the right choice for better results and to speed up customer query processing. In this paper we addressed the issues related to agile method and how the DevOps resolved the problems of agile.

REFERENCES RÉFÉRENCES REFERENCIAS

1. Top Technologies For The Infrastructure & Operations BT Agenda, February 4, 2015, Laura Koetzle, Glenn O'Donnell, Eveline Oehrlich with Christopher Voce, Michael Caputo, Megan Doerr.
2. DevOps Market - By Technology (Sonar Qube, Continuous Integration (CI), Lean startup, Agile software development, Automation, Cloud Computing) - Business Models, Technology Roadmap, Advancements, Forecasts & Analysis (2014 - 2019)
3. White paper on What Smart Businesses Know About DevOps, TechInsights Report | September 2013.
4. X. Yuan and A. M. Memon. Iterative execution-feedback model-directed GUI testing. Information & Software Technology, 52(5):559–575, 2010.
5. Tsong Yueh Chen, Fei-Ching Kuo, Zhi Quan Zhou, “Automated software test data generation” Proceedings of the Fifth International Conference on Quality Software (QSIC05) 2005 IEEE.
6. Xiao Qu, Myra B. Cohen, Katherine M. woolf, “Combinatorial Interaction Regression Testing: A study of Test Case Generation and Prioritization” , 2007
7. Rullian Zhao and Qing Li., “Automatic test generation for dynamic data structure” , Technical report, Beijing University of Chemical Technology, 2006
8. Harrold, “Testing: A Roadmap”, ACM Trans. on Software Engineering 2000

Table I : Comparison of Agile and DevOps.

Methodolgy	Principles	Methods	Practices	Tools
Agile	Follows strategic approaches	XP, Scrum	Standups, planning poker, backlogs, CI, all the specific artifacts a developer uses to perform their work.	JIRA Agile aka Greenhopper
DevOps	System thinking, Amplifying feedback loops (Gene Kim), “Infrastructure as code “are commonly cited DevOps principles.	Same as agile but Scrum with operations Kanban with perations and style change control	Automated Testing, Integrated Configuration Management, Integrated Deployment Planning	jenkins, travis, team city puppet, chef, ansible, cfengine, AWS, OpenStack, vagrant, docker





# GLOBAL JOURNALS INC. (US) GUIDELINES HANDBOOK 2015

---

[WWW.GLOBALJOURNALS.ORG](http://WWW.GLOBALJOURNALS.ORG)

## FELLOWS

### FELLOW OF ASSOCIATION OF RESEARCH SOCIETY IN COMPUTING (FARSC)

Global Journals Incorporate (USA) is accredited by Open Association of Research Society (OARS), U.S.A and in turn, awards “FARSC” title to individuals. The 'FARSC' title is accorded to a selected professional after the approval of the Editor-in-Chief/Editorial Board Members/Dean.



- The “FARSC” is a dignified title which is accorded to a person’s name viz. Dr. John E. Hall, Ph.D., FARSC or William Walldroff, M.S., FARSC.

FARSC accrediting is an honor. It authenticates your research activities. After recognition as FARSC, you can add 'FARSC' title with your name as you use this recognition as additional suffix to your status. This will definitely enhance and add more value and repute to your name. You may use it on your professional Counseling Materials such as CV, Resume, and Visiting Card etc.

*The following benefits can be availed by you only for next three years from the date of certification:*



FARSC designated members are entitled to avail a 40% discount while publishing their research papers (of a single author) with Global Journals Incorporation (USA), if the same is accepted by Editorial Board/Peer Reviewers. If you are a main author or co-author in case of multiple authors, you will be entitled to avail discount of 10%.

Once FARSC title is accorded, the Fellow is authorized to organize a symposium/seminar/conference on behalf of Global Journal Incorporation (USA). The Fellow can also participate in conference/seminar/symposium organized by another institution as representative of Global Journal. In both the cases, it is mandatory for him to discuss with us and obtain our consent.



You may join as member of the Editorial Board of Global Journals Incorporation (USA) after successful completion of three years as Fellow and as Peer Reviewer. In addition, it is also desirable that you should organize seminar/symposium/conference at least once.

We shall provide you intimation regarding launching of e-version of journal of your stream time to time. This may be utilized in your library for the enrichment of knowledge of your students as well as it can also be helpful for the concerned faculty members.

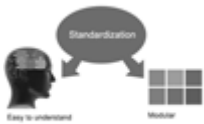




Journals Research  
inducing researches

The FARSC can go through standards of OARS. You can also play vital role if you have any suggestions so that proper amendment can take place to improve the same for the benefit of entire research community.

As FARSC, you will be given a renowned, secure and free professional email address with 100 GB of space e.g. [johnhall@globaljournals.org](mailto:johnhall@globaljournals.org). This will include Webmail, Spam Assassin, Email Forwarders, Auto-Responders, Email Delivery Route tracing, etc.



The FARSC will be eligible for a free application of standardization of their researches. Standardization of research will be subject to acceptability within stipulated norms as the next step after publishing in a journal. We shall depute a team of specialized research professionals who will render their services for elevating your researches to next higher level, which is worldwide open standardization.

The FARSC member can apply for grading and certification of standards of their educational and Institutional Degrees to Open Association of Research, Society U.S.A. Once you are designated as FARSC, you may send us a scanned copy of all of your credentials. OARS will verify, grade and certify them. This will be based on your academic records, quality of research papers published by you, and some more criteria. After certification of all your credentials by OARS, they will be published on your Fellow Profile link on website <https://associationofresearch.org> which will be helpful to upgrade the dignity.



The FARSC members can avail the benefits of free research podcasting in Global Research Radio with their research documents. After publishing the work, (including published elsewhere worldwide with proper authorization) you can upload your research paper with your recorded voice or you can utilize chargeable services of our professional RJs to record your paper in their voice on request.

The FARSC member also entitled to get the benefits of free research podcasting of their research documents through video clips. We can also streamline your conference videos and display your slides/ online slides and online research video clips at reasonable charges, on request.





The FARSC is eligible to earn from sales proceeds of his/her researches/reference/review Books or literature, while publishing with Global Journals. The FARSC can decide whether he/she would like to publish his/her research in a closed manner. In this case, whenever readers purchase that individual research paper for reading, maximum 60% of its profit earned as royalty by Global Journals, will be credited to his/her bank account. The entire entitled amount will be credited to his/her bank account exceeding limit of minimum fixed balance. There is no minimum time limit for collection. The FARSC member can decide its price and we can help in making the right decision.

The FARSC member is eligible to join as a paid peer reviewer at Global Journals Incorporation (USA) and can get remuneration of 15% of author fees, taken from the author of a respective paper. After reviewing 5 or more papers you can request to transfer the amount to your bank account.



## MEMBER OF ASSOCIATION OF RESEARCH SOCIETY IN COMPUTING (MARSC)

The ' MARSC ' title is accorded to a selected professional after the approval of the Editor-in-Chief / Editorial Board Members/Dean.

The "MARSC" is a dignified ornament which is accorded to a person's name viz. Dr. John E. Hall, Ph.D., MARSC or William Walldroff, M.S., MARSC.



MARSC accrediting is an honor. It authenticates your research activities. After becoming MARSC, you can add 'MARSC' title with your name as you use this recognition as additional suffix to your status. This will definitely enhance and add more value and repute to your name. You may use it on your professional Counseling Materials such as CV, Resume, Visiting Card and Name Plate etc.

*The following benefits can be availed by you only for next three years from the date of certification.*



MARSC designated members are entitled to avail a 25% discount while publishing their research papers (of a single author) in Global Journals Inc., if the same is accepted by our Editorial Board and Peer Reviewers. If you are a main author or co-author of a group of authors, you will get discount of 10%.

As MARSC, you will be given a renowned, secure and free professional email address with 30 GB of space e.g. [johnhall@globaljournals.org](mailto:johnhall@globaljournals.org). This will include Webmail, Spam Assassin, Email Forwarders, Auto-Responders, Email Delivery Route tracing, etc.





We shall provide you intimation regarding launching of e-version of journal of your stream time to time. This may be utilized in your library for the enrichment of knowledge of your students as well as it can also be helpful for the concerned faculty members.

The MARSC member can apply for approval, grading and certification of standards of their educational and Institutional Degrees to Open Association of Research, Society U.S.A.



Once you are designated as MARSC, you may send us a scanned copy of all of your credentials. OARS will verify, grade and certify them. This will be based on your academic records, quality of research papers published by you, and some more criteria.

It is mandatory to read all terms and conditions carefully.



## AUXILIARY MEMBERSHIPS

### Institutional Fellow of Open Association of Research Society (USA)-OARS (USA)

Global Journals Incorporation (USA) is accredited by Open Association of Research Society, U.S.A (OARS) and in turn, affiliates research institutions as “Institutional Fellow of Open Association of Research Society” (IFOARS).



The “FARSC” is a dignified title which is accorded to a person’s name viz. Dr. John E. Hall, Ph.D., FARSC or William Walldroff, M.S., FARSC.

The IFOARS institution is entitled to form a Board comprised of one Chairperson and three to five board members preferably from different streams. The Board will be recognized as “Institutional Board of Open Association of Research Society”-(IBOARS).

*The Institute will be entitled to following benefits:*



The IBOARS can initially review research papers of their institute and recommend them to publish with respective journal of Global Journals. It can also review the papers of other institutions after obtaining our consent. The second review will be done by peer reviewer of Global Journals Incorporation (USA) The Board is at liberty to appoint a peer reviewer with the approval of chairperson after consulting us.

The author fees of such paper may be waived off up to 40%.

The Global Journals Incorporation (USA) at its discretion can also refer double blind peer reviewed paper at their end to the board for the verification and to get recommendation for final stage of acceptance of publication.



The IBOARS can organize symposium/seminar/conference in their country on behalf of Global Journals Incorporation (USA)-OARS (USA). The terms and conditions can be discussed separately.

The Board can also play vital role by exploring and giving valuable suggestions regarding the Standards of “Open Association of Research Society, U.S.A (OARS)” so that proper amendment can take place for the benefit of entire research community. We shall provide details of particular standard only on receipt of request from the Board.



The board members can also join us as Individual Fellow with 40% discount on total fees applicable to Individual Fellow. They will be entitled to avail all the benefits as declared. Please visit Individual Fellow-sub menu of GlobalJournals.org to have more relevant details.

We shall provide you intimation regarding launching of e-version of journal of your stream time to time. This may be utilized in your library for the enrichment of knowledge of your students as well as it can also be helpful for the concerned faculty members.



After nomination of your institution as “Institutional Fellow” and constantly functioning successfully for one year, we can consider giving recognition to your institute to function as Regional/Zonal office on our behalf.

The board can also take up the additional allied activities for betterment after our consultation.

### **The following entitlements are applicable to individual Fellows:**

Open Association of Research Society, U.S.A (OARS) By-laws states that an individual Fellow may use the designations as applicable, or the corresponding initials. The Credentials of individual Fellow and Associate designations signify that the individual has gained knowledge of the fundamental concepts. One is magnanimous and proficient in an expertise course covering the professional code of conduct, and follows recognized standards of practice.



Open Association of Research Society (US)/ Global Journals Incorporation (USA), as described in Corporate Statements, are educational, research publishing and professional membership organizations. Achieving our individual Fellow or Associate status is based mainly on meeting stated educational research requirements.

Disbursement of 40% Royalty earned through Global Journals : Researcher = 50%, Peer Reviewer = 37.50%, Institution = 12.50% E.g. Out of 40%, the 20% benefit should be passed on to researcher, 15 % benefit towards remuneration should be given to a reviewer and remaining 5% is to be retained by the institution.



We shall provide print version of 12 issues of any three journals [as per your requirement] out of our 38 journals worth \$ 2376 USD.

### **Other:**

**The individual Fellow and Associate designations accredited by Open Association of Research Society (US) credentials signify guarantees following achievements:**

- The professional accredited with Fellow honor, is entitled to various benefits viz. name, fame, honor, regular flow of income, secured bright future, social status etc.



- In addition to above, if one is single author, then entitled to 40% discount on publishing research paper and can get 10% discount if one is co-author or main author among group of authors.
- The Fellow can organize symposium/seminar/conference on behalf of Global Journals Incorporation (USA) and he/she can also attend the same organized by other institutes on behalf of Global Journals.
- The Fellow can become member of Editorial Board Member after completing 3yrs.
- The Fellow can earn 60% of sales proceeds from the sale of reference/review books/literature/publishing of research paper.
- Fellow can also join as paid peer reviewer and earn 15% remuneration of author charges and can also get an opportunity to join as member of the Editorial Board of Global Journals Incorporation (USA)
- • This individual has learned the basic methods of applying those concepts and techniques to common challenging situations. This individual has further demonstrated an in-depth understanding of the application of suitable techniques to a particular area of research practice.

**Note :**

“

- In future, if the board feels the necessity to change any board member, the same can be done with the consent of the chairperson along with anyone board member without our approval.
- In case, the chairperson needs to be replaced then consent of 2/3rd board members are required and they are also required to jointly pass the resolution copy of which should be sent to us. In such case, it will be compulsory to obtain our approval before replacement.
- In case of “Difference of Opinion [if any]” among the Board members, our decision will be final and binding to everyone.

”



## 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 brevity.

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  $1.4 \times 10^{-3} \text{ m}^3$ , or 4 mm somewhat than  $4 \times 10^{-3} \text{ m}$ . 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](http://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](mailto: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](mailto:editor@globaljournals.org).

You must strictly follow above Author Guidelines before submitting your paper or else we will not at all be responsible for any corrections in future in any of the way.



Before start writing a good quality Computer Science Research Paper, let us first understand what is Computer Science Research Paper? So, Computer Science Research Paper is the paper which is written by professionals or scientists who are associated to Computer Science and Information Technology, or doing research study in these areas. If you are novel to this field then you can consult about this field from your supervisor or guide.

#### TECHNIQUES FOR WRITING A GOOD QUALITY RESEARCH PAPER:

**1. Choosing the topic:** In most cases, the topic is searched by the interest of author but it can be also suggested by the guides. You can have several topics and then you can judge that in which topic or subject you are finding yourself most comfortable. This can be done by asking several questions to yourself, like Will I be able to carry our search in this area? Will I find all necessary recourses to accomplish 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 an 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.



## THE ADMINISTRATION RULES

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 perceptives 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
<i>Abstract</i>	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
<i>Introduction</i>	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
<i>Methods and Procedures</i>	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
<i>Result</i>	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
<i>Discussion</i>	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
<i>References</i>	Complete and correct format, well organized	Beside the point, Incomplete	Wrong format and structuring





# INDEX

---

---

## ***D***

Dehzangi · 5

---

## ***E***

Eigenmaps · 24

---

## ***H***

Hypermy · 26, 27

---

## ***J***

Jenkins · 30, 31

---

## ***K***

Karypis · 24

---

## ***L***

Laplacian · 24

---

## ***P***

Portioning · 2

Procmal · 30

---

## ***S***

Shehata · 26

---

## ***T***

Tenenbaum · 26



save our planet



# Global Journal of Computer Science and Technology

Visit us on the Web at [www.GlobalJournals.org](http://www.GlobalJournals.org) | [www.ComputerResearch.org](http://www.ComputerResearch.org)  
or email us at [helpdesk@globaljournals.org](mailto:helpdesk@globaljournals.org)



ISSN 9754350