Survey on Clustering Techniques in Data Mining for Software Engineering

Maninderjit Kaur[#1], Sushil Kumar Garg[*2]

[#]*Research Scholar (Department of Computer Science and Engineering), RIMT-Institute of Engineering & Technology, Mandi Gobindgarh, Fatehgarh Sahib, Punjab, India*

[1]maninderjit91@yahoo.com

[*]*Principal RIMT-Maharaja Aggrasen Engineering College, Mandi Gobindgarh, Fatehgarh Sahib, Punjab, India*

[2]sushilgarg70@yahoo.com

***Abstract*— Quality and reliability of the computer software is very important. Software development uses a huge amount of software engineering data. Software Engineering data is the collection of execution traces, code bases, graphs, bug reports etc. Software Engineering data is very useful in understanding the development and working of any product or software. Software is of high quality and highly reliable if it is error-free. Software is error-free if there is no bug present in it or it is free from bugs. Bugs are very hard to find. Software Engineering tasks are Programming, Testing, Bug Detection, Debugging and Maintenance. Data Mining Techniques are applied on software engineering tasks. Data mining techniques are used to mine software engineering data and extract the meaningful and useful information. Techniques used for mining software engineering data are matching, clustering, classification etc.**

***Keywords*— Software Engineering, Data Mining, Software Engineering Data, Software Engineering Task, Clustering, Clustering Techniques**

I. INTRODUCTION

The advancement in technology is increasing day by day. This advancement in technology affects the working of different products or softwares. To support these some changes or manipulations become necessity. Maintenance of softwares is becoming very difficult and challenging task. 60% of total life cycle efforts spent on maintenance activities only as in [21]. Software is highly productive and reliable if the Programming, Bug Detection, Testing, Debugging and Maintenance tasks are good as in [12].

Software Engineering Researchers are not expert to develop a tool or algorithm for data mining. In the same way, Data Mining Researchers do not understand the mining requirements in software engineering domain. There is a need of a close collaboration between both domains so the software engineering tasks like Programming, Bug Detection, Testing, Debugging and Maintenance improved. Software engineering data is available in the form of documentation, source code, bug databases, mailing history, bug reports execution traces and graphs as in [2].

Data Mining is the process of finding a small set of precious information and patterns from large sets of raw material. Human are better at storing data. Extracting knowledge from these large datasets is not done in a better way by humans. It is not easy to understand large datasets and finding out the valuable and accurate information to create a good software. Data Mining Process's steps are- data integration, data cleaning, data selection, data transformation, data mining, pattern evaluation and knowledge presentation. Software Engineering data is present in vast amount. Different type of users requires different type of data. All the data is not meaningful for all the users of software. Every user requires the data that is meaningful to them. Meaningful data can be extracted by using different data mining process as in [28]. The data mining process is shown following in Fig.1.

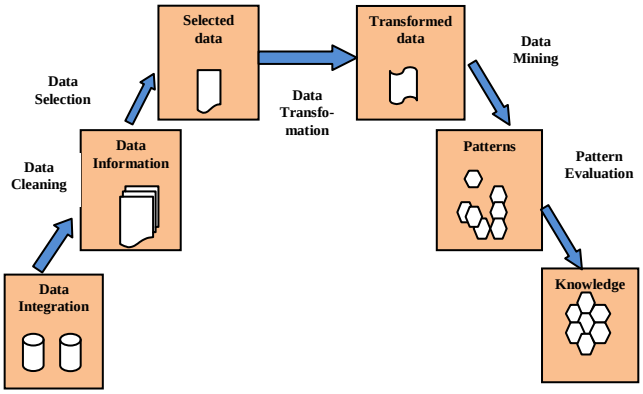

Fig. 1 Data Mining Process [28]

In [27], Mining algorithms fall into four main categories-

- Frequent Pattern Mining: In this, commonly occurring patterns are found.
- Pattern Matching: In this, data instances for given patterns are found.
- Clustering: In this, grouping of data into different clusters is done.
- Classification: In this, predicting labels of data based on already labelled data is done.

In [2], Software Engineering data falls under three categories-

- Sequences: Static execution traces extracted from source code and dynamic execution traces extracted at run time.
- Text: Documentation, source code, e-mails, bug reports, code comments and bug databases.

- Graphs: Static call graphs extracted from source code and dynamic call graphs collected at run time.

Clustering has been applied in many fields like computer science, engineering, social sciences, economics, earth sciences, life and medical sciences [24].

This paper is organized in VI sections. Section II describes the literature review, Section III explains clustering. Section IV gives the overview of types of clusters. Section V explains Clustering techniques. Finally, section VI describes the conclusion.

II. LITERATURE REVIEW

In literature review, we discuss some previous work done in the field of software engineering using data mining. Several conclusions are made with regard to the fitting of clustering techniques.

Suma. V et al. [26] brought out the empirical analysis of several projects developed at various software industries having different production capabilities. They considered defect count to be one of the influencing parameters to predict the success of the projects. Various clustering algorithms are applied on the empirical projects. The observational inferences show that K-Means is more efficient than other algorithms in terms of processing time, efficiency and scalability. A. V. Krishna Prasad et al. [2] design and implement a source code management program. This program scans the code, outputs code to slightly different format. It improves the quality of application. This routine parses tokens from an ANSI C++ file, format the file, extract the header files and colorize a file using data mining techniques. Deqing Wang et al. [8] propose a tool Rebug-Detector. This tool is used to detect the related bugs using bug information and code features. They evaluate the tool on an open source project: Apache LUCENE-Java. This tool is useful to find real bugs in existing projects. Wahidah Husain et al. [28] discuss the techniques for mining software engineering data. They suggest data mining techniques that are used to solve problems in each type of software engineering data. As a result suggest FIM (Frequent Itemset Mining) and FSM (Frequent Sequence Mining) appropriate for mining sequence data, Classification is best for mining graph data and Text mining is best for mining text data. Kapila Kapoor and Geetika Kapoor [12] propose CLEMENTE tool that mines useful patterns out of scattered data. This tool is having many applications in different areas like public sector, drug discovery, bioinformatics, web mining and customer relationship management.

III. CLUSTERING

Clustering is known by different names in different areas, such as numerical taxonomy (ecology, biology), unsupervised learning (pattern recognition), partition (graph theory) and typology (social sciences). According to definition, "Cluster Analysis is the art of finding groups in data" as in [15]. In [3], [10], Clustering is a machine learning technique, in which set of data objects are grouped into multiple groups or clusters. Similarity between the objects of same cluster (intra-class) is more than the similarity between objects of other clusters (inter-class). Data objects' group having some common features are called clusters. Attribute's characteristics are used to calculate similarities and dissimilarities between the objects. Cluster should have two main properties-

- High intra-class similarity
- Low inter-class similarity

Clustering is an unsupervised learning technique. There is no labelled data is available [24]. It is the form of learning by observation rather than learning by example. Clustering is used to allot labels to unlabelled data. No pre-existing grouping is known for unlabelled data [10], [6]. In [17], [6] the advantages and disadvantages of clustering are-

- **Advantages**: Without user intervention provide automatic recovery from failure. Clustering provides incremental growth to group new data as use of personal computers and mobile technology is increased.
- **Disadvantages**: In case of database corruption unable to recover and complexity.

A. Clustering Process

Input to clustering process is the real data that is dirty. The number of groups/clusters forming a partition is the output of clustering process. Pre-processing is done on raw data to prepare it for clustering. Data cleaning, remove noise and inconsistent data. Data integration, multiple data sources may be combined. Data transformation, data are transformed into forms appropriate for mining. Data reduction, the volume of the representation is reduced but result remains same. The clustering process is shown in Fig.2.

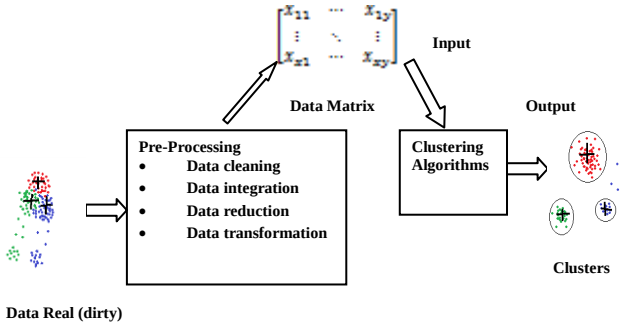

Fig. 2 Clustering Process [13]

In mathematical form the goal of clustering is described as follows-

$$X = C_1 \cup \ldots\ldots\ldots C_i \cup C_n; \quad C_i \cap C_j = \emptyset \ (i \neq j) \qquad [13]$$

Where X denotes original data set, n is the no. of clusters and C_i,C_j are clusters of X as in [13].

IV. TYPES OF CLUSTERS

In clustering, five types of clusters exist on the basis of their characteristics. These are Center-based clusters, Density-based clusters, well-separated clusters, Contiguous clusters and Shared Property or Conceptual clusters as in [4], [5], [9], [19], [6].

A. *Center-Based Clusters*

A cluster is a set of objects. An object in cluster is more closer (similar) to the "center" of a cluster, not to the center of any other cluste. A centroid (average of all points in cluster) or a mediod (most representative point in cluster) is often the center of a cluster.

Fig. 3 Center-Based Clusters [11]

B. *Density-Based Clusters*

A cluster is separated by low-density regions, from other regions of high-density. A cluster is a dense region of points. When noise and outliers are present and when clusters are irregular, this definition is more often used.

Fig. 4 Density-Based Clusters (6 Density-Based Clusters) [11]

C. *Well-Separated Clusters*

A cluster is a set of nodes such that any node in a cluster is closer or more similar to every other node in the same cluster than to any node not in the cluster. Sometimes threshold can be used to specify similarity or closeness between the nodes in cluster.

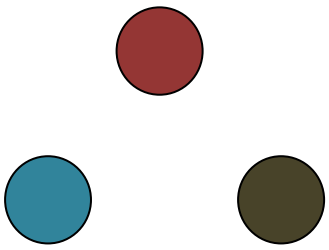

Fig. 5 Well-Separated Cluster [11]

D. *Contiguous Clusters (Nearest Neighbour or Transitive)*

A cluster is set of points such that a point in a cluster is closer to one or more other points in the cluster than to any point not in the cluster.

Fig. 6 Contiguous Clusters (8 Contiguous Cluster) [11]

E. *Shared Property or Conceptual Clusters*

Find clusters that share some common property or represent a particular concept.

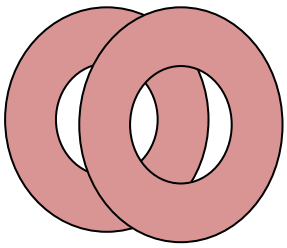

Fig. 7 Conceptual Cluster (2 Overlapping Circles) [11]

V. CLUSTERING TECHNIQUES

In [4] Clustering Techniques are different from one another on the basis of-

- Manner of clustering (whether the object belongs to one cluster or more than one).
- Use of threshold values in the construction of clusters.
- Similarity measuring procedures.

Clustering can be hard partitioned or soft partitioned. In hard partitioning, each node belongs to only one cluster. It is based on the strict logic [24]. In soft partitioning, a node may also be belongs to all clusters. Membership degree of a node is given for detail see [15]. Clustering Algorithms can use the concept of different techniques. A clustering algorithm can belongs to more than one technique. Clustering techniques can be categorized into partition-based, hierarchical-based, density-based, grid-based and many other techniques as in [4].

A. *Partition-Based Clustering Techniques*

Partition-Based Clustering Techniques [9] [20], [24], [4], [5], [23], [17], creates one level (un-nested) partitioning. In this, data points 'n' are split into 'K' partitions on the basis of certain criterion function or objective function. Minimizing square error is one such objective function that is most widely used and is computed as-

$$E = \sum\sum \| p - m_i \|^2 \quad [4]$$

Where m_i is mean of cluster, p is a point in a cluster. For reassigning points between K clusters iteratively it uses relocation schemes. Each partition must contain at least one object and each object must belong to exactly one partition. In single pass partition method there are three steps. First is to make centroid, second is calculate similarity and final is to compare similarity with specified threshold and assign objects to cluster. Very efficient for serial processor but first cluster formed may be larger than other created later.

- **Advantages:** Partitioned-based techniques are easy to implement. Used for large data sets.
- **Disadvantages:** Partitioned-based techniques are sensitive to noisy data. Number of clusters and stopping criteria are user defined. This technique produces spherical shaped clusters. It gives poor result

due to overlapping of data points, when a point is closer to the center of another cluster.

There are many algorithms that use this technique for clustering. These are K-Mean, Bisecting, K-Mean, Medoids Method, PAM(Partitioning Around Medoids), CLARA (Clustering Large Applications), Probabilistic Clustering, FCM (Fuzzy C-Means), FFT (Farthest First Traversal K-Center), CLARANS, ISODATA, Fuzzy K-Means, K-Modes, Fuzzy K-Modes, K-Prototype. K-Mean algorithm is the most well known and simple algorithm. K-Mean algorithm is the standard algorithm. It has linear time complexity. Several variants of K-Mean algorithm exists that reduces the problems exist in K-Mean. [4], [9], [5], [3], [23], [6], [7], [15], [22], [11], [19], [16], [18], [25], [14], [24] contain in depth analysis of these algorithms.

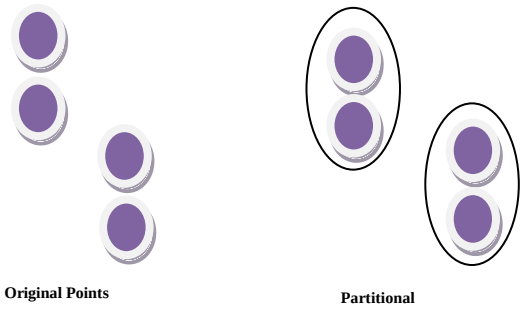

Fig. 7 Partitional Clustering [5]

B. *Hierarchical Clustering Technique*

Hierarchical Clustering [4], [20], [24], [5], [6], [9], [13], [3], [17], [1], [19], [18] constructing a hierarchy of clusters by dividing the data set. Construction of clusters is step by step. This technique takes series of partition not a single step partition. This hierarchy graphically represented by a diagram called a Dendrogram or by binary tree. Root node represents whole data set and leaf represents a data object. Height of dendrogram represents distance between an object and a cluster, between each pair of an object or between each pair of clusters. These are of two types-

1) Agglomerative Nesting (AGNES): Bottom-up approach for constructing the hierarchy of clusters. This technique merges or combines the 'n' objects into groups. Merging operation is based on certain criteria like maximum distance, minimum distance, average distance, center distance. Algorithms that use this technique are Single Linkage (Nearest neighbour), Complete Linkage, Group Average Linkage, Median Linkage, Centroid linkage as in [7]. For detail see [9], [13].

2) *Divisive Analysis (DIANA):* Top-down approach for constructing hierarchy of clusters. This technique separate 'n' objects into groups. It starts from the root step by step each node form leaf. This is not widely used method. For 'n' objects cluster there is '2^n-1' two-subset divisions are possible, which is very expensive. Algorithms of this technique are MONA (Monothetic Analysis), Bisecting K-Mean method.

- **Advantages:** This technique can easily handle any form of similarity or distance. Applicable to all types of attributes, flexible, easy to implement and gives best results in some cases. There is no need to specify the number of clusters by user and produce better quality clusters.
- **Disadvantages:** This technique is sensitive to noise and outliers, not capable of correcting previous misclassification. Vagueness of termination criteria. Different distance measures generate different results. It is very difficult to identify correct number of clusters.

New hierarchical clustering algorithms are CURE, ROCK, CHAMELEON, SLINK, COBWEB, CLINK, LEGCLUST, BIRCH, RHC (Relative Hierarchical Clustering). For detail see [13], [24], [22]. Hierarchical clustering is a theoretical foundation of cluster analysis but considered obsolete as in [7]. In [13], AGNES and DIANA applied to document clustering, DIANA used in linguistics, information retrieval, document clustering applications.

TABLE I
LINKAGE METHOD

Single Linkage	$d_{12} = \min_{ij} d(X_i, Y_j)$	Single Linkage is defined as minimum distance between elements of each cluster.
Complete Linkage	$d_{12} = \max_{ij} d(X_i, Y_j)$	Complete Linkage is maximum distance between elements of each Cluster.
Average Linkage	$d_{12} = \frac{1}{kl} \sum_{i=1}^{k} \sum_{j=1}^{l} d(X_i, Y_j)$	Average Linkage is mean distance between elements of each cluster.

C. *Density-Based Partitioning Techniques*

Density-based partitioning techniques [4], [5], [20], [22], [23] are one-scan technique. It finds clusters according to the regions which grow with high density. Clusters are high density area than remaining data set. Density is the number of objects in a cluster. It finds arbitrary shaped clusters. It is applicable to spatial data. This technique is of two types-

1) Based on density function.

2) Based on connectivity of points.

- **Advantages:** It does not require any specified number of clusters by user. It can handle noise.
- **Disadvantages:** In case of neck type of data sets it fails. It does not work well in case of high dimensionality data.

There are many algorithms that fall under this category; these are DBSCAN, GDBSCAN, OPTICS, DBCLASD, DENCLUE, and SNN. For detail see [4], [3], [5], [20], [7], [18], [24], [22].

D. Grid-Based Partitioning Technique

Grid-based partitioning technique [20], [4], [22], [23], [11], [18], [5], [19], uses multidimensional grid data structure. It does not concern with data points. It deals with the value space that surrounds the data points. To form clusters uses dense grids and multi-resolution grid data structure. It quantized the original data space into finite number of cells which form the grid structure and then perform all the operations on the quantized space. It focuses on spatial data. It maps infinite number of data records to finite number of grids. User defined grid size and density thresholds. Adaptive grids overcome this problem.

- **Advantages:** Fastest processing time depends on size of grid. It can work with attributes of various types.
- **Disadvantages:** Due to mesh size computation load for clustering increases.

Algorithms of this technique are STING, Wave Cluster, CLIQUE, OPTICS, FC (Fractal Clustering). [3], [11], [18], [22] contain in depth detail about algorithms. No distance computation is done in this, which is the major feature. Representation of clusters is done in a more meaningful way. Steps followed by any grid-based algorithm is-

1. Divide the data space into finite number of cells.
2. For each cell, calculate cell density.
3. According to their densities sort cells.
4. Identifying clusters centers.
5. Traversal of neighbour cells.

Perform well over time complexity and high dimensional data among various clustering algorithms.

E. Other Techniques

There are many other techniques for clustering exists like Graph theory based clustering, Kernel based clustering, Search technique based clustering, Subspace clustering, Neural network based clustering, Constraint based clustering, Model based clustering. Some of them are explained as follows-

1) *Graph Theory Based Clustering*: In Graph theory based clustering [24], [11], concepts of graph theory are used. Nodes (V) of a weighted graph (G) correspond to data points in pattern space and edges (E) reflect the proximities between each pair of data points. It is very simple and scalable. It is very important in VLSI designs. It is of two types-

- Between-graph: It divides set of graphs into different clusters.
- Within-graph: It divides nodes of a graph into different clusters.

Algorithms of this clustering technique are Single Linkage, Complete Linkage, CLICK, CHAMELEON, and CAST.

2) Neural Network Based Clustering: Neural network based technique is dominated by ART (Adaptive Resonance Theory) and SOFMs. For Detail see [24], [1].

3) *Subspace Clustering*: In [20], Subspace clustering work with high dimensional data. It uses the subspace of actual dimension. It uses the idea of grid-based and density-based clustering techniques. Algorithms that use this technique are CLIQUE, ENCLUS, PROCLUS, ORCLUS, and MAFIA. For detail see [18].

4) Model-Based Clustering: Model-based clustering identifies cluster with a certain model whose unknown parameters have to be found. Algorithms under this category are EM, SNOB, AUTOCLASS, MCLUST, K-Means, K-Medoids, ISODATA, Forgy, Bisecting K-Means, x-Means, Kernel K-Means, and SOM as in [20], [23].

VI. CONCLUSIONS

In this survey paper, we provide the discussion of data mining for software engineering. We also provide discussion about the clustering techniques. Data mining is most efficient technique to manage large amount of data since information is highly valuable and expensive. Every technique has to solve different problems and have their own advantages and disadvantages. There is no such clustering technique and algorithm exists that is used to solve all the problems and is a best fit for all applications. As the application changes requirements will also change. With this change the selection of clustering technique affected. No technique or algorithm is the readymade solution to all applications and problems. Predefined number of clusters and stopping criteria affect the accuracy and performance of clustering. Handling of noisy data, data set size, shape of the clusters all affects the clustering results. Based on the application, we have to choose the suitable clustering technique and algorithm in future work.

ACKNOWLEDGMENT

The author would like to thank the RIMT Institutes, Mandi Gobindgarh-147301, Fatehgarh Sahib, Punjab, India. Author would also wish to thank editors and reviewers for their valuable suggestions and constructive comments that help in bringing out the useful information and improve the content of paper.

REFERENCES

[1] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data Clustering: A Review," ACM Computing Surveys, vol. 31, No. 3, pp. 264-323, September 1999.

[2] A. V. Krishna Prasad, and Dr. S. Rama Krishna, "Data Mining for Secure Software Engineering- Source Code Management Tool Case Study," International Journal of Engineering Science and Technology (ISSN: 0975-5462), vol. 2(7), 2010, 2667-2677.

[3] Aastha Joshi, and Rajneet Kaur, "A Review: Comparative Study of Various Clustering Techniques in Data Mining," IJARCSSE:

International Journal of Advanced Research in Computer Science and Software Engineering (ISSN: 2277 128X), vol. 3, Issue 3, March 2013.

[4] Amandeep Kaur Mann, and Navneet Kaur, "Review Paper on Clustering Techniques," Global Journal of Computer Science and Technology (ISSN (Online): 0975-4172), vol. 13, Issue 5, Version 1.0, 2013.

[5] Amandeep Kaur Mann, and Navneet Kaur, "Survey Paper on Clustering Techniques," IJSETR: International Journal of Science, Engineering and Technology Research (ISSN: 2278-7798), vol. 2, Issue 4, April 2013.

[6] Amanpreet Kaur Toor, and Amarpreet Singh, "A Survey Paper on Recent Clustering Approaches in Data Mining," IJARCSSE: International Journal of Advanced Research in Computer Science and Software Engineering (ISSN: 2277 128X), vol. 3, Issue 11, November 2013.

[7] Anoop Kumar Jain, and Satyam Maheswari, "Survey of Recent Clustering Techniques in Data Mining," International Archive of Applied Sciences and Technology (ISSN: 0976-4828), vol. 3[2], pp. 68-75, June 2012.

[8] Deqing Wang, Mengxiang Lin, Hui Zhang, and Hongping Hu, "Detect Related Bugs from Source Code Using Bug Information", 34th Annual IEEE Computer Software and Applications Conference, 2010.

[9] Er. Arpit Gupta, Er. Ankit Gupta and Er. Amit Mishra, "Research Paper on Cluster Techniques of Data Variations," IJATER: International Journal of Advanced Technology Engineering Research (ISSN: 2250-3536), vol. 1, Issue 1, November 2011.

[10] Jiawei Han, Micheline Kamber and Jian Pei, *Data Mining: Concepts and Techniques*, 3rd ed., 2013.

[11] K. Kameshwaran, and K. Malarvizhi, "Survey on Clustering Techniques in Data Mining," IJCSIT: International Journal of Computer Science and Information Technologies (ISSN: 0975-9646), vol. 5(2), pp. 2272-2276, 2014.

[12] Kapila Kapoor and Geetika Kapoor, "Improving Software Reliability and Productivity through Data Mining", Proceedings of the 5th National conference; INDIACom-2011, March 10-11, 2011.

[13] M. Kuchaki Rafsanjani, Z. Asghari Varzaneh, and N. Emami Chukanlo, "A survey of hierarchical clustering algorithms," TJMCS: The Journal of Mathematics and Computer Science, vol. 5, No. 3, pp. 229-240, December 2012.

[14] Manpreet Kaur, and Usvir Kaur, "A Survey on Clustering Principles with K-Means clustering Algorithms Using Different Methods in Detail," International Journal of Computer Science and Mobile Computing, vol. 2, Issue 5, pg. 327-331, May 2013.

[15] Mrutyunjaya Panda, and Manas Ranjan Patra, "Some Clustering Algorithms to Enhance the Performance of the Network Intrusion Detection System," *JATIT: Journal of Theoretical and Applied Information Technology*, 2008.

[16] Ms. Asmita Yadav, "A Survey of Issues and Challenges Associated with Clustering Algorithms," IJSETT: International Journal for Science and Emerging Technologies with Latest Trends (ISSN (Online): 2250-3641), vol. 10(1), pp. 7-11, 2013.

[17] P. IndiraPriya, and Dr. D. K. Ghosh, "A Survey on Different Clustering Algorithms in Data Mining Technique," IJMER: International Journal of Modern Engineering Research (ISSN: 2249-6645), vol. 3, Issue 1, pp. 267-274, Jan-Feb. 2013.

[18] Pavel Berkhin, "A Survey of Clustering Data Mining Techniques", *Accure Software, Inc.*

[19] Pradeep Rai, and Shubha Singh, "A Survey of Clustering Techniques," International Journal of Computer Applications (ISSN: 0975-8887), vol. 7, No. 12, October 2010.

[20] Prof. Neha Soni, and Prof. Amit Ganatra, "Categorization of Several Clustering Algorithms from Different Perspective: A Review," IJARCSSE: International Journal of Advanced Research in Computer Science and Software Engineering (ISSN: 2277 128X), vol. 2, Issue 8, August 2012.

[21] Rajib Mall, *Fundamentals of Software Engineering*, 3rd ed., 2010.

[22] Rama. B, Jayashree. P, and Salim Jiwani, "A Survey on clustering," IJCSE: International Journal on Computer Science and Engineering (ISSN: 0975-3397), vol. 02, No. 09, pp. 2976-2980, 2010.

[23] Ramandeep Kaur, and Dr. Gurjit Singh Bhathal, "A Survey of Clustering Techniques," IJARCSSE: International Journal of Advanced Research in Computer Science and Software Engineering (ISSN: 2277 128X), vol. 3, Issue 5, May 2013.

[24] Rui Xu, and Donald Wunch II, "Survey of Clustering Algorithms", IEEE Transaction on Neural Networks, vol. 16, no. 3, May 2005.

[25] S. Revathi, and Dr. T. Nalini, "Performance Comparison of Various Clustering Algorithm," IJARCSSE: International Journal of Advanced Research in Computer Science and Software Engineering (ISSN: 2277 128X), vol. 3, Issue 2, February 2013.

[26] Suma. V, Pushpavathi t.P, and Ramaswamy. V, "An Approach to Predict Software Project Success by Data Mining Clustering", international Conference on Data Mining and Computer Engineering (ICDMCE'2012), Bangkok (Thailand), December 21-22, 2012.

[27] V. Neelima, Annapurna. N, V. alekhya, and Dr. B. M. Vidyavathi, "Bug Detection through Text Data Mining," IJARCSSE: International Journal of Advanced Research in Computer Science and Software Engineering (ISSN: 2277 128X), vol. 3, Issue 5, pp. 564–569, May 2013.

[28] Wahidah Husain, Pey Ven Low, Lee Koon Ng, and Zhen Li Ong, "Application of Data Mining Techniques for Improving Software Engineering", ICIT 2011 the 5th International Conference on Information Technology.

