Segmentation and Clustering of Stakeholders using Data Mining Techniques

Prerana Chaithra[1*], ShantharamNaya[1], Nanda Kumar A. N[2]

[1] RVCE, Bangalore

[2] CSE, GSSSIETW, Mysore

* Corresponding author email: preranachaithra15@gmail.com

Abstract

Stakeholders have lot of influence on the software product. Hence their involvement and influence during the product development is high. There are different types of stakeholders who are either directly or indirectly involved in the development of a software. Having the awareness of these stakeholders software requirements will help the developers to develop a robust software product. Therefore, in this paper we are presenting the different techniques that help the developers to segregate the different types of stakeholders and the major roles played by them in the production of the software product.

Index Terms- Stakeholders, software product, software requirements, Data Mining

1   INTRODUCTION

Stakeholder involvement is crucial for the successful development of the products in the IT industry. Stakeholders must be involved in the software development lifecycle as shown in figure 1. Among the stakeholders some are Active Stakeholders who are directly involved in the production of the software product. Especially in Agile method where the development must be done at a faster rate, stakeholders play a major role. The development team must understand the needs or the requirements of the stakeholder correctly in order to develop the software product fastly.
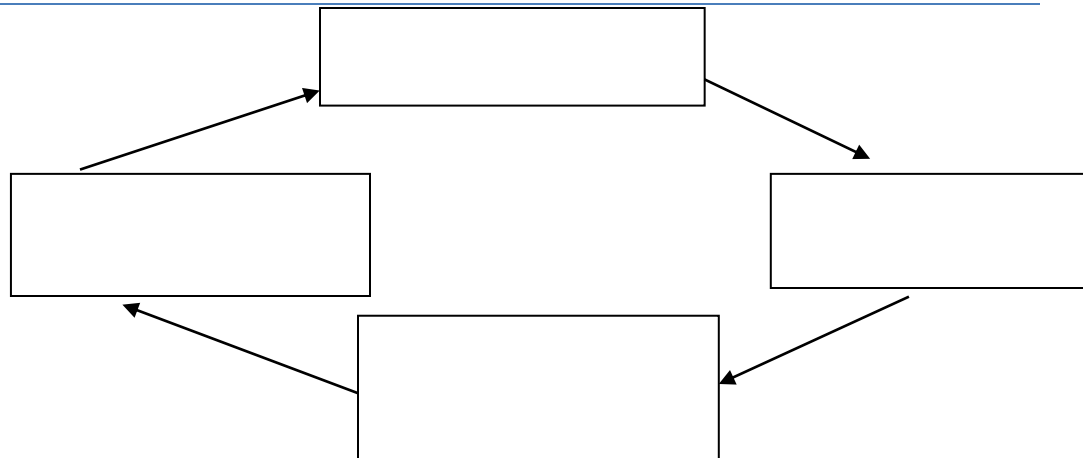
Figure 1 : Involvement of stakeholders in the software development life cycle

2   LITERATURE SURVEY

2.1   Role of the stakeholders

Stakeholders play a major role during software requirement analysis. Their inputs are crucial during the software analysis phase. Their involvement during the requirements phase of the software product development will help the Business analysts to come up with better requirements for the product development. They help to phrase requirements in a quick and easy manner.

In the design phase if the designer communicates with the stakeholders, it will be easier for the designer to design the product according to the needs of the stakeholders [1]. Software architect discusses the architecture with the main stakeholders and come up with the tailor made architecture for the software product.

Constant touch the prime stakeholder helps in the faster implementation of the software product. Implementation will be robust and the product will adhere to its architecture. Incremental development needs constant interactions with the stakeholder [2]. Agile method of software development revolves around the stakeholder for quick and robust development. Acceptance testing needs the user who is also a stakeholder to be mainly involved in testing the software product. The satisfaction of the stakeholders will assure that the software product is according to the software requirements given by the stakeholders. If the product passes the acceptance testing, it is a clear indication that the product can be deployed in its designated place [3]-[5].

Involvement of the stakeholders in the software development lifecycle helps to produce a better and more satisfiable software product. It helps to prevent unnecessary loss of resources during the software development.

Proceedings of the 3rd National Conference on Image Processing, Computing, Communication, Networking and Data Analytics (NCICCNDA 2018)

483

2.2 Identification of Stakeholders

One of the important criteria is to identify the stakeholders for that product. It is difficult to identify the correct stakeholders. Some smart steps must be considered to identify them [5].. Stakeholders can be identified by the answering the following questions:

- ➢ Who requested for the product?
- ➢ Who is financing the product?
- ➢ Who will provide supplemental project funding, if needed?
- ➢ Who will be at loss if the correct product is not developed?
- ➢ Who will make profit if the product is completed successfully?
- ➢ Who will give the requirements for the quality of the product?
- ➢ Which company will supply the resources for the product.?
- ➢ Who is providing technological and/or software resources for this project?
- ➢ Who is providing knowledge resources or special expertise for this project?
- ➢ Who is providing the human resource for this product?
- ➢ Who will take the responsibility if the product is not according to its requirements?
- ➢ Who can trigger the development of the product?

2.3 Types of stakeholders and their influence

Various types of stakeholders who influence the software product development are as follows

- ➢ User of the product
- ➢ Customer
- ➢ Developing organization
- ➢ Engineers of the customer organization
- ➢ Customer (person or organization)
- ➢ Management of the developing organization
- ➢ Management of the customer organization
- ➢ Technical experts
- ➢ Marketing personnel
- ➢ Analysts
- ➢ Designers
- ➢ Government entities
- ➢ Development team
- ➢ Testing team
- ➢ Project managers
- ➢ Resource managers
- ➢ Tech lead
- ➢ Vendors
- ➢ Partners
- ➢ Government entities

➢ Advertisers
➢ Collaborating companies
➢ Procuring managers
➢ Law of the land providers

3 DESIGN AND IMPLEMENTATION

In our paper we are considering the following stakeholders
- Customer
- End user
- Project manager
- Resource managers
- Development team
- Testing team
- Government entities

3.1 Segmentation Techniques

For segmenting the different types of stakeholders and their influences we are using Histogram and Linear Regression techniques of Data Mining.

3.2 Histogram technique

Histogram technique is one of the best ways to give the summary of data. It considers the records in the database for the information [5]. Table 1 gives the summary of database to create a histogram of the more influential stakeholder based on the involvement in the product development by counting the number of occurrences of different influences in our database. For this example database of 7 records are considered. The results provide the influences of various stakeholders on the software product.

The histogram technique holds good for database with few records but for a database of large number of records we need better techniques. One such technique is Linear Regression technique which will be discussed later on in the paper. Figure 3 shows the number of stakeholders with various percentage of influence on the product.

Table 1: Database of Stakeholders with different influence types

ID	Name	Importance	Influence	Profit
1	Customer	High	On product	Good product
2	End User	Very High	On product	Good product
3	Project Manager	Mediocre	On technology used	Company goals
4	Resource Manager	High	On resource used	Company goals
5	Development Team	Very High	On Implementation	Develop better product
6	Testing Team	Mediocre	On quality	New testing Technique
7	Government Entities	Low	On regulations	Enforce rules

Proceedings of the 3ʳᵈ National Conference on Image Processing, Computing, Communication, Networking and Data Analytics (NCICCNDA 2018)

485

3.3 Linear Regression Technique

Regression technique is used for statistical prediction. There are a different types of regressions which are used in statistics. Linear regression technique is one such technique. It helps to reduce the occurrence of errors. It consists of predictor and prediction. A model is developed that maps values from predictors to the prediction which results in reduction of errors [6]. The relationship between the predictors and the predictions can be mapped on a two dimensional graph [11]. The records are plotted with the predictor values on the X axis and the prediction values on the Y axis of the graph.

The linear regression model can be viewed as the line graph that reduced the error rate between the actual prediction value and the value on the prediction graph from the linear regression model. This graph is depicted in the Figure 4. The linear regression model builds a predictive model that is a line that maps between a predictor value to a prediction value. Among the various possible lines that are drawn through the data the one that minimizes the distance between the line and the data points is the one that is selected for the predictive model.

3.4 Clustering Techniques

Clustering techniques are of two types, those that produce a hierarchy of clusters and those that do not produce the hierarchy of clusters. The hierarchical clustering techniques produce a hierarchy of clusters from small to the big. Clustering is a technique that has no absolute value for the answer. Therefore depending on the concerned application of the clustering, number of clusters may vary according to the need. Hierarchy of clusters lead to the selection of clusters that are required. It is possible to have as many clusters as the number of records present in the database [2]. In our case the records within the cluster are relatively similar to each other and certainly different from the other clusters. There may be fewer clusters than the number of records in the database. Formation of the number of clusters depends on the interpretation and anticipation. Hierarchical clustering method is more advantageous as it allows the end user to select from various clusters or from minimal number of clusters.

Hierarchy of clusters can be shown as a tree where the smallest clusters of stakeholders are merged together to create the next highest level of clusters of stakeholders and those at that level merge together to create the next highest level of clusters. Creation of such hierarchy of clusters lead to the determination of the right number of clusters. This cautiously summarizes the data when providing useful information. Figure 2 shows the hierarchy of clusters. Clusters of stakeholders at the lowest level are merged together to form larger clusters of stakeholders at the next level of the hierarchy.
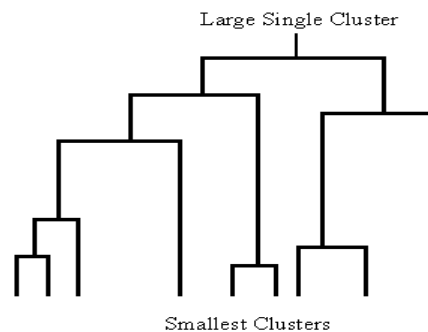

Figure 2 Hierarchy clustering for stakeholders

The above technique can be used to agglomerate the influence of the stake holders on the product. So the hierarchy of the stakeholders can be formulated as follows. End User, Resource Manager, Project Manager, Customer, End User, Resource Manager, Project Manager, Customer, Development Team, Testing Team and Government Entities

4   RESULTS

The results of the segmentation and clustering are as follows:

Result of Histogram techniques is shown in the Figure 3. It shows the number of stakeholders with various percentage of influence on the product.
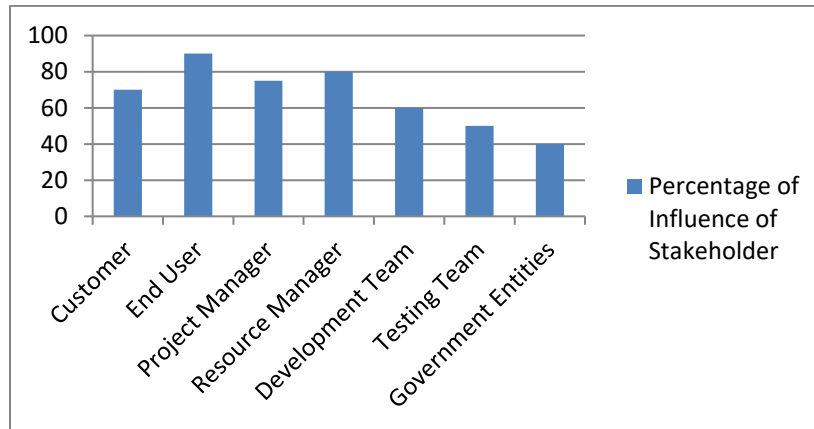
Figure 3: This histogram shows the number of stakeholders with various percentage of influence on the product

Result of Linear Regression technique is shown in the Figure 4. We can conclude that we have to consider the stakeholders in the priority basis of their influence on the product.
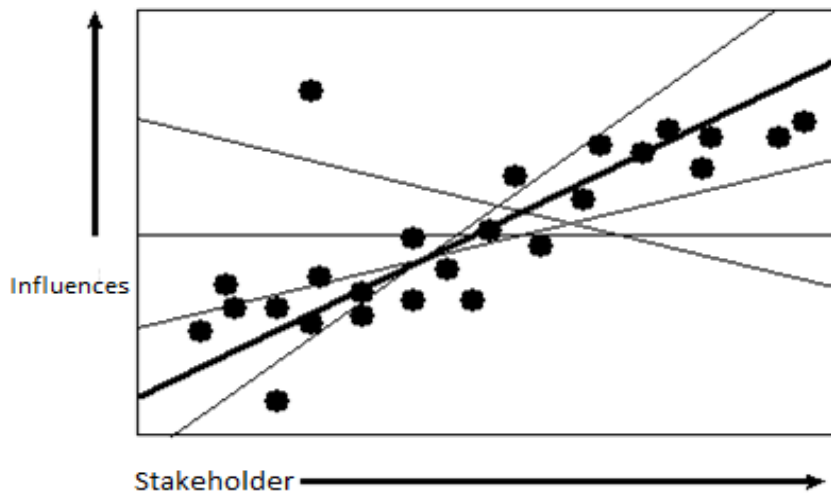
Figure 4 Linear regression for stakeholders and their influence on the product development

Proceedings of the 3rd National Conference on Image Processing, Computing, Communication, Networking and Data Analytics (NCICCNDA 2018)

487

The result of hierarchy of clusters technique can be used to agglomerate the influence of the stake holders on the product. So the hierarchy of the stakeholders can be formulated as follows.

1. End User
2. Customer
3. Resource Manager
4. Project Manager
5. Resource Manager
6. Project Manager
7. Development Team
8. Testing Team
9. Government Entities

5   CONCLUSIONS

We would like to conclude that the above data mining techniques can be used effectively in order to segment the stakeholders in a better way. For smaller products Histogram technique of segmentation is very helpful to the developers. Using Linear regression segmentation and Hierarchical clustering it is very effective to do the segmentation and clustering of the stakeholders. These Data mining techniques are a blessing to the developers to develop the software product effectively and help to release the product based on time to market.

References

[1] C. C. Aggarwal, Y. Zhao, P. S. Yu. On Text Clustering with Side Information, ICDE Conference, 2012.

[2] C. C. Aggarwal, P. S. Yu. A Framework for Clustering Massive Text and Categorical Data Streams, SIAM Conference on Data Mining, 2006.

[3] C. C. Aggarwal, S. C. Gates, P. S. Yu. On Using Partial Supervision for Text Categorization, IEEE Transactions on Knowledge and Data Engineering, 16(2), 245–255, 2004.

[4] P. Andritsos, P. Tsaparas, R. Miller, K. Sevcik. LIMBO: Scalable Clustering of Categorical Data. EDBT Conference, 2004.

[5] www.thearlin.com/text/dmtechniques/dmtechniques.htm

[6] C. C. Aggarwal, P. S. Yu. On Effective Conceptual Indexing and Similarity Search in Text, ICDM Conference, 2001.

[7] C. C. Aggarwal, P. S. Yu. Finding Generalized Projected Clusters in High Dimensional Spaces, ACM SIGMOD Conference, 2000.

[8] R. Agrawal, J. Gehrke, P. Raghavan. D. Gunopulos. Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications, ACM SIGMOD Conference, 1999.

[9] C. C. Aggarwal, C. Procopiuc, J. Wolf, P. S. Yu, J.-S.Park.Fast Algorithms for Projected Clustering, ACM SIGMOD Conference, 1999.

[10] J. Allan, R. Papka, V. Lavrenko. Online new event detection and tracking.ACM SIGIR Conference, 1998.

[11] R. Agrawal, R. Srikant. Fast Algorithms for Mining Association Rules in Large Databases, VLDB Conference, 1994.