IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 08, Issue 4 (April. 2018), ||VI|| PP 84-88

Implementation of Pam Cluster for Evaluating SaaS on the Cloud Computing Environment

Dhanamma Jagli, Dr. (Mrs.) Seema Purohit, Dr.N. Subhash Chandra
[1]Research Scholar and Assistant Professor, VESIT, [2, 3]Research Guide, JNTU Hyderabad.
Corresponding Auther:Dhanamma Jagli

ABSTRACTION: Cloud Computing has emerged as a new paradigm in the field of network-based services within many industries and application domainsand a dynamic computing sharing resources. Cloud computing paradigm has emerged and that is transforming the it industry at huge. In cloud computing, all resources are available as services and accessible through the internet. Especially Software-As-A-Service (SaaS) is service delivery model that support end users to access any software or an application as a service via the internet without installing at locally. The usage of saas has been increased by many users and thus leads to need to evaluate the quality of saas to select the best one that suits to cloud services users. In this paper, a quality model is implemented by using Data MiningPartitioning Around Medoids (Pam) clustering model for evaluating the quality of software as a service (SAAS) in the cloud computing environment.

KEYWORDS: Software As A Service (Saas),Cloud Computing, *Partitioning* Around Medoids(PAM).

Date of Submission: 22-03-2018 Date of acceptance: 07-04-2018

I. INTRODUCTION

Cloud Computing Is a Tremendous Resource Sharing Computing Adopted By Many Organizations inthe Last Decade. The Main Concept of Cloud Computing Has Used Any Resource as A Service. I.E Everything as A Service (Xaas).they are mainly three service models: IAAS (infrastructure as a service), PAAS (platform as a service) and SAAS (software as a service).the SAAS is deployment model where the service end user need not install the software on their local machine even though they use it as it's locally. The SaaS have been using by many vendors as well as providing by many cloud service providers because of its advantages. However, the usage of the SaaS has been increased throughout the globe in the computing world. Hence,many challenges were introduced, one of the main challenges for cloud service users is how to select right service as per their requirements and which one would meet their expectations. In order to deal with this particular challenge, in this paper is a new quality model is proposed to evaluate the qualityof SaaS in the cloud computing environment based on the data mining clustering algorithm.This paper initially describes the importance of data mining pam-clustering algorithm- , then it explains about research methodology adopted for specified challenge. Finally, it explains about partitioning around medoids (PAM) clustering implementation using r-studio with SaaS quality related data.

II. LITERATURE REVIEW

Jerry gao, pushkala pattabhiraman, Xiaoping bai w. T. Tsai presented their research work [7] as new formal graphic models and metrics to evaluate SaaS performance and scalability features. The results shown best potential application and effectiveness of the proposed model for evaluating SaaS scalability and performance attributes only. But not on other attributes, which are also playing an important role for good quality. Zia urRahmanproposed work [8] discussed and proposed a multi-criteria cloud service selection methodology in general. Very important parameters like reliability, trust, reputation, etc. Are not given importance even though they are very critical in the cloud computing environment. Qiang he, Jun Han, its proposed work [9] is used to evaluate the attribute multi-tenancy cloud-based software applications with less scalability.

It may not suitable if number of end users are increasing. Tung-hsiang Chou andwan-ting liu research work [10] presentedthat some of theSaaS dimensions, integrated along with service dimensions of servequl to maintain the standard for customer's service. So that presented work is only benefited with very few attributes of SaaS, not applied to quality parameters. From the literature review, it has been identified that in order to evaluate SaaS, quality six attributes are playing a vital role in evaluating the quality like availability, pay for use, data managed by the provider, scalability, reusability, and service customizability as shown in the figure1.

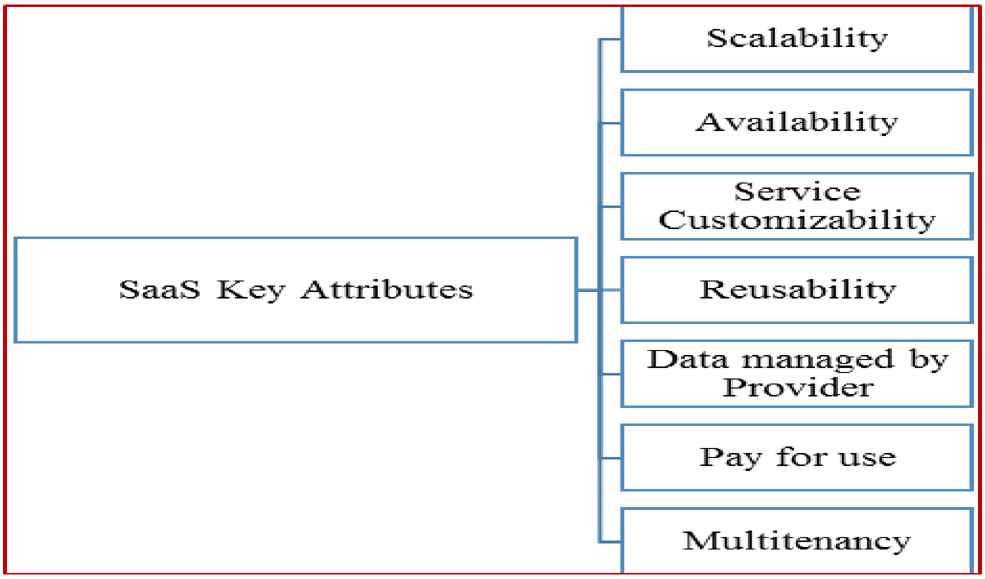

Figure 1: *SaaS Kay Attributes*

III. RESEARCH METHODOLOGY

The clustering is one of the important data mining methods for discovering knowledge in multidimensional data. The goal of clustering is to identify patterns or groups of similar objects within a data set of interest. In the literature, it is referred as “pattern recognition” or “unsupervised machine learning”. Clustering Analysis onaSaaSDataset. Data Preprocessing Steps Have Been Applied Before Applying Clustering Analysis. Steps For Cluster Analysis Are As Follows:

1. Assessing Clustering Tendency.
2. Defining the Optimal Number of Clusters.
3. Computing Clustering Algorithms.
4. Validating Clustering Analysis.

The research methodology is the systematic, theoretical analysis of the methods applied to a field of study. It comprises the theoretical analysis of the body of methods and principles associated with a branch of knowledge.The research methodology has been proposed for a given problem and proposed as shown in the below figure 2.

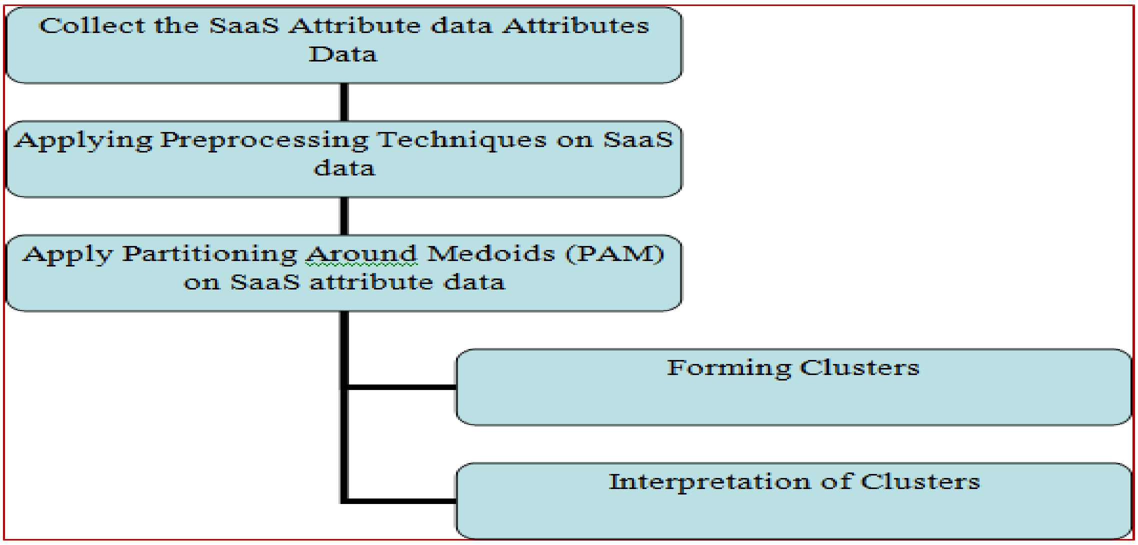
Collect the SaaS Attribute data Attributes Data
Applying Preprocessing Techniques on SaaS data
Apply Partitioning Around Medoids (PAM) on SaaS attribute data
Forming Clusters
Interpretation of Clusters

Figure 2: The Work Flow of Methodology

IV. PARTITIONING AROUND MEDOIDS(PAM)

Partitioning aroundMedio's(PAM) Clustering Forms Clusters through Defining the Use of Means Implies That K-Means Clustering Is Highly Sensitive to Outliers. A more robust algorithm is provided by partitioning around medio's (pam) algorithm which is also known as k-medio's clustering.The PAM algorithm is based on the search for k representative objects or medio's among the observations of the dataset. The Goal Is To Find K-Representative Objects Which Minimize The Sum Of The Dissimilarities Of The Observations Of Their Closest Representative Object.

PAM Algorithm

- Build Phase:

1. Select k objects to become the medio's, or in case these objects were provided use them as the medio's;
2. Calculate the dissimilarity matrix if it was not provided;
3. Assign every object to its closest mediod;

- Swap Phase:

1. For each cluster search if any of the objects of the cluster decrease the average dissimilarity coefficient; if it does, select the entity that decreases this coefficient the most as the mediod for this cluster;
2. If at least one mediod has changed go to (3), else end the algorithm.

V. IMPLEMENTATION IN R

About The Data

The SaaS Data Set Has Been Collected From Amazon Public Cloud And The Data Set Has The Following Attributes:

✓ Product name: the name the SaaS product.
✓ Rating: rating is a numeric value given by SaaS users on the scale of 1 to 10, 1 is lowest, 10 is highest.
✓ No. Of reviews: reviews is numeric attribute, no. Of users given a comment on that particular product
✓ Category of SaaS: this is a categorical attribute, describes the category of SaaS.

The data have been analyzed in the r studio. A clustering analysis is performed with more details. Start with defining the optimal number of clusters using elbow method. This method is to validate the number of clusters and the idea of the elbow method is to run k-medio's clustering of the dataset for a range of values of k.

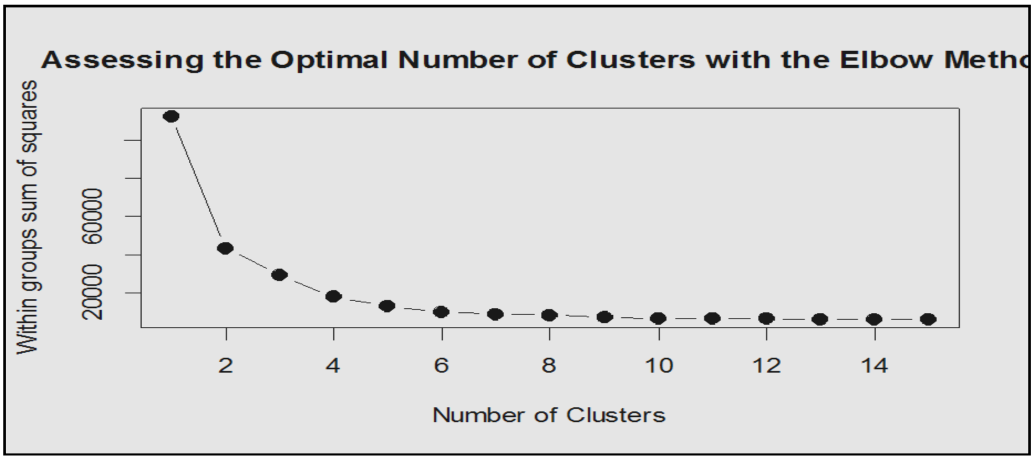

Figure 4:*Optimal Clusters forPAM*

It has been analyzed by using elbow method that after 4 no changes in the curve so the optimal number of clusters the elbow method suggests is =4.

Advantages:

✓ K-mediod's method is more robust than k-means in the presence of noise and outliers
✓ Pick actual objects to represent clusters instead of mean values.

Disadvantages:

✓ K-mediod's is more costly that the k-means method.
✓ Like k-means, k-medio's requires the user to specify k.
✓ It does not scale well for large data sets.

Required R Function andPackages

- ✓ The function pam(), which simplified format is:
- ✓ Pam(x, k, metric = "euclidean", stand = false)
- ✓ X: possible values includes:
- ✓ K: the number of clusters
- ✓ Metric: the distance metrics to be used. Available options are “euclidean” and “manhattan”.
- ✓ Stand: logical value; if true, the variables (columns) in x are standardized before calculating the dissimilarities. Ignored when x is a dissimilarity matrix.

- ✓ The function pam(), which simplified format is:
- ✓ Pam(x, k, metric = "euclidean", stand = false)
- ✓ X: possible values includes:
- ✓ K: the number of clusters
- ✓ Metric: the distance metrics to be used. Available options are “euclidean” and “manhattan”.
- ✓ Stand: logical value; if true, the variables (columns) in x are standardized before calculating the dissimilarities. Ignored when x is a dissimilarity matrix.

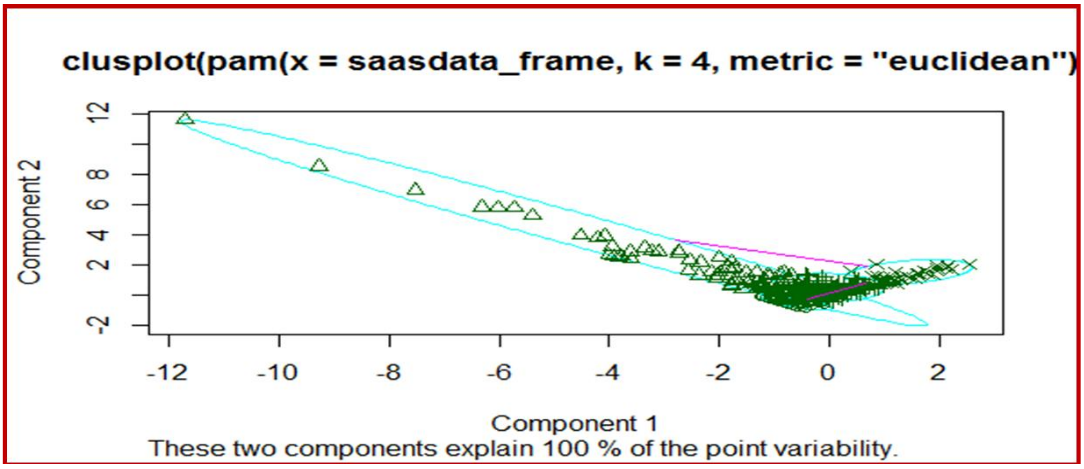

Figure 4: Cluster Formation Using PAM Cluster

Desirable Cluster Characteristics

- ✓ Within cluster dissimilarity should be small.
- ✓ Between clusters dissimilarity should be large.
- ✓ Members should be well represented by its centroid.
- ✓ Clusters should be stable.
- ✓ All clusters should have roughly the same size.
- ✓ Dunn index should be high
- ✓ The average distance within cluster to be as small.
- ✓ The average distance between clusters to be as large.
- ✓ Separation of a point in the cluster to a point of another cluster should be high.

Cluster Validations

- External clustering validation: which consists in comparing the results of a cluster analysis to an externally known result.
- Internal clustering validation: which use the internal information of the clustering process to evaluate the goodness of a clustering structure.
- Relative clustering validation: which evaluates the clustering structure of varying different parameter values for the same algorithm.
- The two commonly used indices for assessing the goodness of clustering: the silhouette width and the dunn index
- If the data set contains compact and well-separated clusters, the diameter of the clusters is expected to be small. It is proved for pam clusters.
- The distance between the clusters is expected to be large and is true for pam clusters.
- Thus, dunn index should be maximized for good cluster formation and is proved by using pam clusters.

VI. CONCLUSION

In this paper, the quality of any SaaS product with identified quality attributes have been proposed by using model based partitioning around medio's (pam) clustering algorithm. It is also implemented in the R studio statically tool. Partitioning Around Medio's (PAM) clustering is good for k-medio's method is more robust than k-means in the presence of noise and outliers. In the future, it has been proposed to analyse for large data sets.

REFERENCES

[1] N. Alluring, A. Smith, And D. Turnbull, "Clustering With EM And K-Means," Univ. San Diego, California, Pp. 261–95, 2003.

[2] A. KakaAnd A. Kaka, "Expectation Maximization Tutorial Expectation-Maximization Algorithm For Clustering Multidimensional Numerical Data Expectation Maximization Tutorial," Vol. 2012, No. November, 2014.

[3] T. M. Mitchell, "Expectation Maximization, And Learning fromPartly Unobserved Data (Part 2)," Mach. Learn., Vol. 15, No. April, 2005.

[4] S. Bormann, "The Expectation Maximization Algorithm A Short Tutorial," Submit. Publ., Vol. 25, No. X, Pp. 1–9, 2009.

[5] E. M. MeandP. Abele, "Maximum Likelihood (ML), Expectation Maximization (EM)," No. Ml, Pp. 1–23.

[6] H. Borland, Y. Koenig, andN. Morgan, "REMAP: Recursive Estimation and Maximization of A Posteriori Probabilities in Connectionist Speech Recognition." Euro speech, 1995.

[7] S. Bormann, "The Expectation Maximization Algorithm A Short Tutorial," Submit. Publ., Vol. 25, No. X, Pp. 1–9, 2009.

[8] C. B. Do And S. Batzoglou, "What Is The Expectation Maximization Algorithm?" Nat. Biotechnol., Vol. 26, No. 8, Pp. 897–899, 2008.

[9] Dhanamma Jagli, S. Purohit, And N. S. Chandra, "SAASQUAL: A Quality Model For Evaluating Saas On The Cloud Computing Environment," 2015.

[10] Gupta Jagli, Mrs. Dhanamma, ".Clustering Model For Evaluating Saas," 2013.

[11] Dhanamma Jagli, Dr.Sunita Mahajan, Dr.Subhash Chandra"CBC Approach For Evaluating Potential Saas On The Cloud," Vesit.Edu, Vol. 2, Pp. 43–49, 2014.

[12] J. Y. Lee, J. W. Lee, D. W. Cheun, And S. D. Kim, "A Quality Model For Evaluating Software-As-A-Service In Cloud Computing," In Software Engineering Research, Management And Applications, 2009. SERA '09. 7th ACIS International Conference On, 2009, Pp. 261–266.

[13] Ibrahim LF. Using Of Clustering Algorithm CWSP-PAM For Rural Network Planning. Proc - 3rd Int Conf Inf Technol Appl ICITA 2005. 2005; I: 280-283. Doi:10.1109/ICITA.2005.300.

[14] Liu D, Graham J. Simple Measures Of Individual Cluster-Membership Certainty For Hard Partitional Clustering. 2017:1-9.

[15] Bott R, Huynh V-N, Pham SB, Et Al. An End-To-End Qos Mapping Approach For Cloud Service Selection. Antonopoulos N, Gillam L, Eds. IEEE Ninth World Congr Serv. 2013; 5(1):341-348. Doi:10.1007/S13398-014-0173-7.2.

[16] Clusters "Around Medoids", A More Robust Version Of K-Means. Usage.

Dhanamma Jagli "Implementation of Pam Cluster for Evaluating Saas on The Cloud Computing Environment "IOSR Journal of Engineering (IOSRJEN), vol. 8, no. 4, 2018, pp. 84-88