

A Hybrid Clustering Model to Evaluate Quality of Software Services on Cloud Environment

Dhanamma Jagli¹, N. Subhash Chandra², Seema Purohit³

¹Research Scholar, JNTU /Hyderabad, India.

Email: dsjagli.vesit@gmail.com.

²Professor, CVR College of Engineering/ CSE Department, Hyderabad, India.

Email: subhashchandra.n.cse@gmail.com

³Professor, Kirti Engineering College/ CSE Department, University of Mumbai, India.

Email: supurohit@gmail.com.

Abstract: Cloud computing is emerging as an innovative paradigm in the field of Internet-based services in many software and hardware industries. The benefits that it provides in terms of cost, flexibility, efficient technology and business ability represent a huge competitive advantage for an organization. This is proven to be an essential requirement for extending several existing applications. In order to utilize the entire rewards of the SaaS model proficiently and successfully, there should be a suitable quality model to evaluate the SaaS quality. In fact, service-providers must evaluate their services against needs of service-user to increase their services. The existing SaaS quality evaluation models lack focus on all aspects of quality and services together.

On the other hand, each cloud service user is unique that leads to a very huge variation in the requirements of the software services. Therefore, this paper presents a hybrid model to help service users to opt a better SaaS product to satisfy their environment and needs. The proposed Hybrid model is implemented by using a hybrid clustering algorithm to construct clusters of SaaS products.

Index Terms: Software as a Service (SaaS), K-Means Clustering, DBSCAN Clustering, Hybrid Clustering

I. INTRODUCTION

Cloud computing is the newest technology of computing as well as to provide numerous scalable resources dynamically and to virtualizes those resources as a service over the Internet. Three main cloud offering models are available in the cloud computing environment, they are Platform as a service (PaaS), Infrastructure as a Service (IaaS), and Software as a Service (SaaS). The SaaS model provides several applications for costumers without installing any application locally. So, one can use any virtual servers available on the web to make a cloud. SaaS is a model of delivering into centrally hosted applications of the Internet. SaaS applications run on a SaaS supplier's server. The advantage of SaaS is that the software or application can be accessed through the internet without installing them at the local system. The cloud service providers manage access to the application, including security, availability, and performance. SaaS business applications are regularly accessed by service users by means of a thin client via a web browser¹. The private and public organizations and individuals are adopting SaaS products to utilize its profit. But, selection of right SaaS product from pool of products is

a major problem for SaaS customer. In this context, a hybrid clustering model has been proposed to improve the quality of evaluation of SaaS products. The paper is organized as follows: Section II presents the literature review carried out towards SaaS product selection and identifies the gaps embedded in the existing models. Section III explains proposed hybrid model, Section IV discusses experimental results and compares with the results of existing models and Section V presents conclusion and future work.

II. LITERATURE REVIEW

The work on SaaS evaluation is relatively very less till 2005; later, analytical works have improved towards SaaS analysis. Chen Yiming, et.al, proposed a model by choosing the most effective SaaS vendor supported service principles instead of selecting SaaS Product². Pang XiongPilar cyst, Li Dong, projected an advanced quality model that measures the safety, software quality of the SaaS and quality of service from the perspective of a service-provider^{3,4,5}. Jae Yoo, et.al, proposed a top-quality methodology that may facilitate to look at the quality side of SaaS, supported the SaaS key features derived from primary SaaS options^{4, 6}. They valued and assessed their proposed quality model for evaluating SaaS quality. However, they have not covered all the key features of SaaS like multi-tenancy etc. Manish Godse, Shrikant Maulik, proposed a replacement model based on Analytic Hierarchy Process (AHP) approaches / approached for ranking the SaaS features like Functionality, Vendor name, Design, Usability, and Cost⁷. However, the proper ranking of SaaS products is not possible, as each feature plays a vital role in the SaaS. Authors Qian Tao etc. proposed a technique for all cloud services based on non-functional QoS attributes like reputation, reliability, security, time, price and availability^{8,9,10}. They exploited the results by Partitioning around Medians (PAM) rule for all kinds of cloud services. However, its own limitations will effect the optimum solution⁹. The review of literature summarizes that, the researchers contributed their innovations and supported key options and quality attributes to evaluate SaaS products. But, all researchers considered vendor perspective attributes only. This paper addresses the gaps involved in evaluation of SaaS without considering user perspective attributes.

III. PROPOSED HYBRID CLUSTERING MODEL

To overcome the research gaps identified from literature review, a Hybrid clustering model is proposed and implemented on real time data sets with user's perspective attributes i.e.

- Product Name: the name with the SaaS product.
- Rating: Rating is a numeric value given by SaaS users on the scale of 1 to 10, 1 being the lowest and 10 highest.
- No. of Reviews: Reviews are a numeric attribute, which give no. of comments from SaaS users of that product.
- Category of SaaS: This is a categorical attribute that describes the category of SaaS.

This SaaS data set has been collected from Amazon public cloud and a few other sources. The proposed hybrid model consists of two phases. In the first phase K- means clustering algorithm applied to SaaS products data set. In the second phase, the DBSCAN algorithm applied to first phase results. The data flow in Hybrid Clustering model is as shown in Figure1.

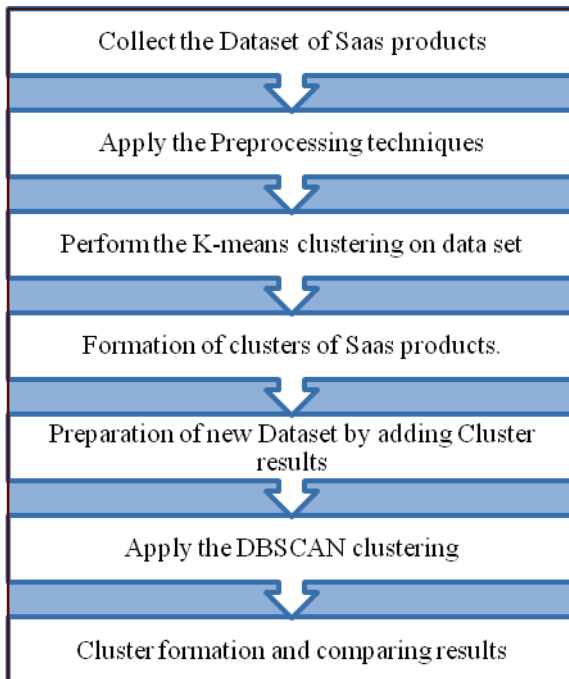


Figure 1. The Work flow of the Proposed Model

K-means Clustering

K-means clustering is “to classify objects of multiple groups such that, objects in the same cluster are as similar as possible, whereas objects of different clusters are as dissimilar as possible”¹¹. In k-means clustering, every cluster is represented by means of its center or centroid that corresponds to the mean of data objects assigned to each cluster. K-means clustering model constructs four clusters of SaaS data set ratings and number of reviews. Different clusters are shown in Figure 2 and indicated in different colors.

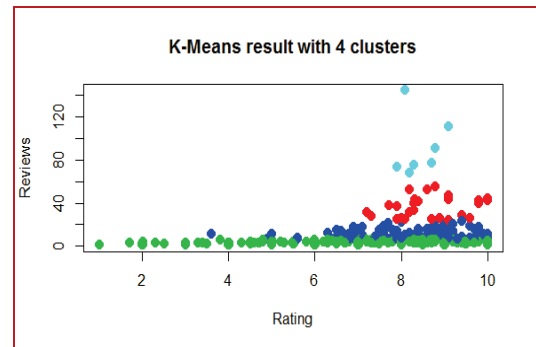


Figure 2. K-means Cluster plot

Cluster 1 represents SaaS products of the highest rating, and highest number of reviews. • Cluster 2 represents SaaS products of the highest rating, and average number of reviews. • Cluster 3 represents SaaS products of average rating and average reviews. • Cluster 4 represents SaaS products that are below average rating, with a below average number of reviews.

Limitations of K-Means:

The limitations of K-means clustering algorithm is as follows: It can be applied only when the mean of a cluster is defined. User needs to specify k. The results obtained are sensitive to the initial random selection of cluster-centers. That means clustering is highly sensitive to outliers. K-means is not suitable for discovering clusters with any convex shapes or clusters of very different size. Hence, the Density Based Clustering Algorithm is implemented over K-means clusters data set, which does not require the user to specify the number of clusters to be generated. It can find any shape of clusters that can identify outliers.

Density Based Clustering (DBSCAN)

DBSCAN is a Density-Based Spatial Clustering and Application with Noise. It clusters together points that are close to each other based on spatial distance and minimum number of points^{12,13}. This cluster model identifies clusters of any shape in a data set and handles noise and outliers within the data set. The steps involved in DBSCAN clustering algorithm are as shown in the Figure 3

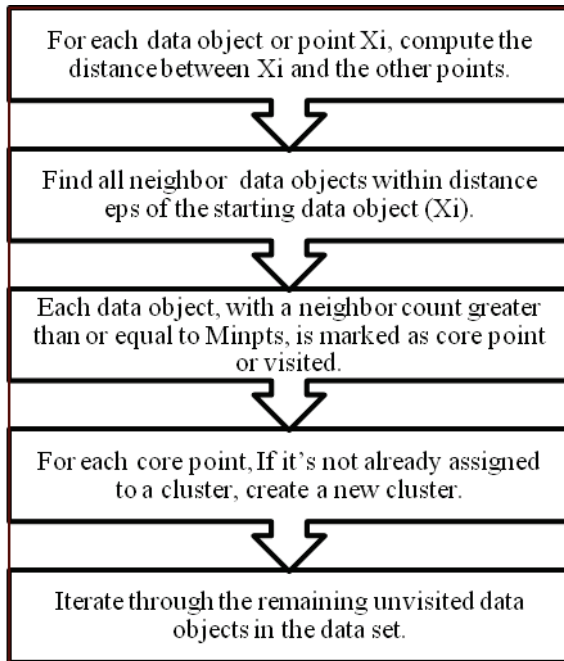


Figure 3. DBSCAN Algorithm

The DBSCAN algorithm requires 2 parameters^{14,15} i.e. Eps: the minimum spatial distance between two points. If the distance between two points is less than or equal to this value (eps), then these points are considered neighbors. MinPoints: the minimum number of points to form a cluster. In conventional DBSCAN Algorithm, Eps and MinPts values were chosen on trial and error basis. To enhance the output of DBSCAN algorithm, the optimal Eps value has been generated by KNNdist in R Studio as 0.15 for a sample dataset, shown in Figure 4.

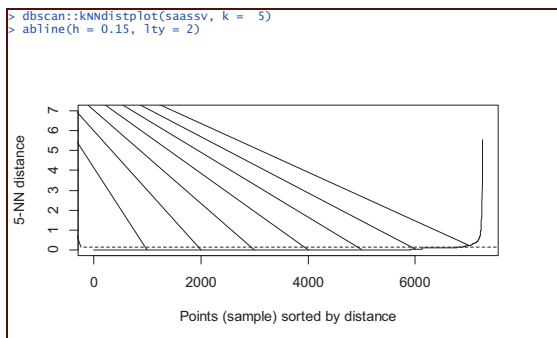


Figure 4. Optimal Eps value for DBSCAN Clustering

DBSCAN algorithm constructed seven different clusters of convex shape as shown in the Figure 5 and represented by different colors. The number of clusters varies according to datasets and it also impacts on no. of noise points. Taking this as a consideration, a cluster number as seven to reduce no. of noise points.

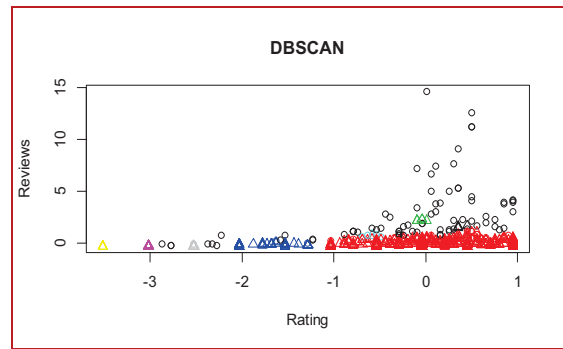


Figure 5. DBSCAN Cluster Plot

Limitations of DBSCAN:

DBSCAN fails to discover the clusters of varying densities and overlapping regions. It is not completely deterministic, so that, border points or data objects that are reachable from more than one cluster can be part of either cluster, depending on the order for data processing.

Hybrid K-means -DBSCAN (KD) Clustering

Each of the K-means and DBSCAN algorithms have their own limitations and benefits. To make use of the advantages of both clustering algorithms, hybrid clustering algorithm has been proposed to overcome the limitations of both clustering algorithms. The steps involved in the proposed Hybrid algorithm are as shown in Figure 6.

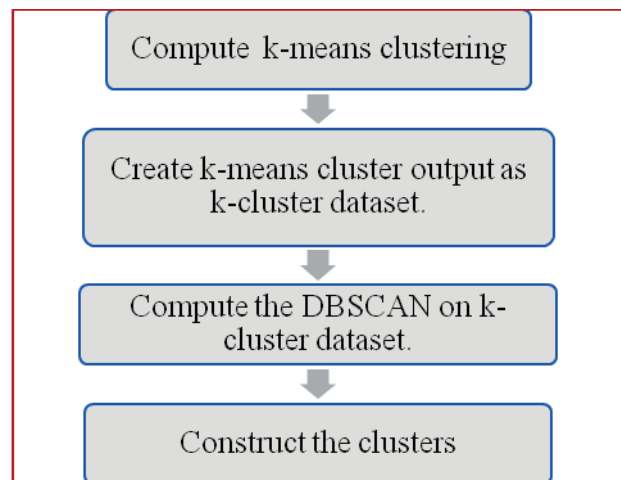


Figure 6. Proposed Hybrid KD Clustering Algorithm

IV. RESULTS AND DISCUSSION

The Hybrid clustering algorithm has constructed 40 clusters and 277 noise points with Eps = 0.15 and minPts = 5 for 1460 objects. The clusters are shown in Figure 7 with different colors.

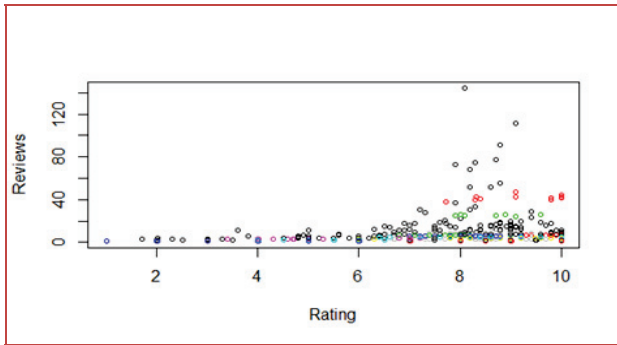


Figure 7. Hybrid KD Cluster Plots

Silhouette value is considered as a performance measure to compare the results of K-means, DBSCAN and Hybrid Algorithms^{16,17}. It indicates that, the similarity of an object is in its own cluster (cohesion) compared to alternative clusters (separation). The silhouette ranges from -1 to $+1$, the higher value indicates that the object is well suited to its own cluster and poorly matched to neighboring clusters. As per results, the proposed Hybrid clustering algorithm has produced comparatively better results than k-means and DBSCAN as shown in the Figure 8.

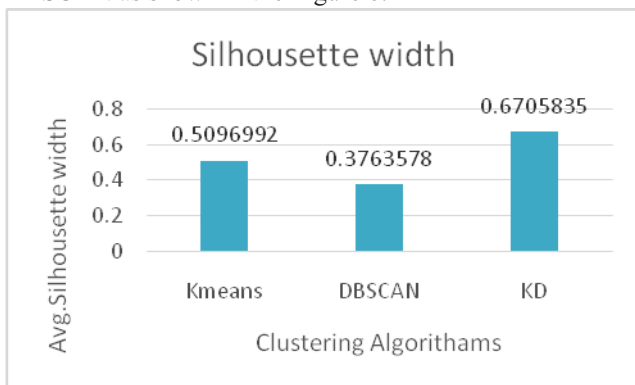


Figure 8. Results Comparison table

V.CONCLUSIONS

In summary, the proposed Hybrid model is one of the cloud offerings, Software as a Service (SaaS). This model provides a clear overview of different parameters of choices of SaaS consumers that can help to evaluate before using SaaS products. This proposed hybrid cluster model is considered attributes of software service from a user's perspective service that is feedback from service users in terms of ratings and reviews. This model provides SaaS product cluster groups so that service users can select any SaaS product of the group accordingly. Moreover, this model is implemented on live data set and produced better results compared to existing models.

REFERENCES

- [1] Benlian A, Koufaris M, Hess T. Service Quality in Software-as-a-Service: Developing the SaaS-Qual Measure and Examining Its Role in Usage Continuance. *J Manag Inf Syst.* 2012;28(3):85-126. doi:10.2753/MIS0742-1222280303.
- [2] Goode S, Lin C, Tsai JC, Jiang JJ. Rethinking the role of security in client satisfaction with Software-as-a-Service (SaaS) providers. *Decis Support Syst.* 2015;70:73-85. doi:10.1016/j.dss.2014.12.005.
- [3] Bott R, Zinn D, Hart Q, et al. The Analytic Hierarchy Process. In: Antonopoulos N, Gillam L, eds. *Igarss 2014. Vol 5. Informatik im Fokus.* IEEE; 2014:1-5. doi:10.1007/s13398-014-0173-7.2.
- [4] 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.
- [5] Wen PX, Dong L. Quality Model for Evaluating SaaS Service. In: 2013 Fourth International Conference on Emerging Intelligent Data and Web Technologies. IEEE; 2013:83-87. doi:10.1109/EIDWT.2013.19.
- [6] Lee JYJW. A Quality Model for Evaluating Software-as-a-Service in Cloud Computing. In: 2009 Seventh ACIS International Conference on Software Engineering Research, Management and Applications. IEEE; 2009:261-266. doi:10.1109/SERA.2009.43.
- [7] Godse M, Mulik S. An approach for selecting Software-as-a-Service (SaaS) product. In: *CLOUD 2009 - 2009 IEEE International Conference on Cloud Computing.* ; 2009:155-158. doi:10.1109/CLOUD.2009.74.
- [8] Tao Q, Chang H, Yi Y, Gu C. A trustworthy management approach for cloud services QoS data. In: *Machine Learning and Cybernetics (ICMLC), 2010 International Conference on.* Vol 4. ; 2010:1626-1631.
- [9] He Q, Han J, Yang Y, Grundy J, Jin H. QoS-driven service selection for multi-tenant SaaS. *Proc - 2012 IEEE 5th Int Conf Cloud Comput CLOUD 2012.* 2012:566-573. doi:10.1109/CLOUD.2012.125.
- [10] Liu F, Guo W, Zhao ZQ, Chou W. SaaS integration for software cloud. *Proc - 2010 IEEE 3rd Int Conf Cloud Comput CLOUD 2010.* 2010:402-409. doi:10.1109/CLOUD.2010.67.
- [11] Jung HW, Kim SG, Chung CS. Measuring software product quality: A survey of ISO/IEC 9126. *IEEE Softw.* 2004;21(5):88-92. doi:10.1109/MS.2004.1331309.
- [12] Desharnais J-M. Analysis of ISO/IEC 9126 and 25010. 2009.
- [13] Pedram H, Moghaddam DK, Asheghi Z. Applying the ISO 9126 model to the evaluation of an e-learning system in Iran. *Iran J Inf Process Manag.* 2012;27(2):495-517.
- [14] Parasuraman A, Berry LL, Zeithaml VA. Refinement and Reassessment of the SERVQUAL Scale. *J Retail.* 1991;67(4):420. doi:10.1111/j.1438-8677.2010.00335.x.
- [15] Parasuraman a, Zeithaml V a, Berry LL. SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality. Vol 64.; 1988:28. doi:10.1016/S0148-2963(99)00084-3.
- [16] Dhanamma Jagli, Seema Purohit S, Chandra NSS, et al. SAASQUAL : A Quality Model For Evaluating SaaS on The Cloud. *Adv Intell Syst Comput.* 2015;654:1-6. doi:10.1007/978-981-10-6620-7_41.
- [17] Dhanamma Jagli M, Purohit S, Chandra NS, et al. SaaS CloudQual: A Quality Model for Evaluating Software as a Service on the Cloud Computing Environment. In: *ICICSE2016.* Springer; 2016:2-5. doi:10.1007/978-981-10-3818-1.