# A Novel Approach for Big Data Analysis using Mobile Cloud Computing

Sonu kumari[1], Paridhi[2]
Research Scholar[1], Asst. Professor[2]
RPS college of engineering & technology , Mahendergarh,
Indira gandhi university,Meerpur Rewari

**Abstract:** The number of mobile subscriptions is steadily rising as mobile computing and wireless networking technologies progress at breakneck speed. For users of mobile devices, mobile cloud apps as well as services are in high demand as a result of this. Thus, mobile cloud computing presents a significant possibility for business and research (MCC). In the proposed research, we are using Mobile Cloud Computing dataset and deploying it as a BigData file on cloud using Databricks and performing feature selection, extraction, correlation, classification and clustering of DataCenter/Virtual Machine based on the Mobile Cloud Data. And all this process is happening at DataBricks Cloud Environment. Machine learning algorithms, KNN and SVM are used for clustering and the better machine learning algorithm is considered.

Keywords: Mobile Cloud Computing; KNN algorithm; SVM algorithm; Big Data; Feature selection; Feature extraction.

## 1. Introduction

Due to the rapid advancement of wireless technology, networking, and mobile computing, mobile subscriptions have recently increased significantly. Between 2011 and 2015, TechNavio forecasts that there will be an increase of 18.12 percent at a CAGR of the North American Enterprise Mobile Cloud Computing market. The increasing need for enterprise mobility is one of the major drivers of this expansion. This market is dominated by the likes of Amazon, IBM, Terremark Worldwide, and Salesforce.com, to name a few.

For mobile users, cloud computing advancements give substantial advantages as cloud infrastructure and platforms enable nearly unlimited processing capacity with scalability and resource sharing capabilities. Many of the current restrictions of mobile computing could be solved with this. These are just a few of the advantages of mobile cloud computing. [3].

Cloud-based services can handle the most demanding workloads and massive amounts of data, allowing mobile devices to use less processing power and storage space.

We can now build more sophisticated mobile applications than ever before since mobile devices have access to a powerful cloud.

There are many ways in which mobile clients might reduce energy consumption without sacrificing performance, such as offloading resource-intensive activity to the cloud.

Using thin mobile clients and a cloud platform, it is possible to design smaller, more efficient mobile devices that outperform their bigger, more powerful equivalents in terms of performance. As a result, mobile clients can be "dumbed down" to merely manage user interaction, with the rest of the application's data and work being transferred to the cloud.

MCC (Mobile Cloud Computing) is a paradigm that aims to take use of cloud computing's advantages and adapt them to the mobile area, only utilising the cloud where the benefits outweigh the costs (i.e. when it lessens resource utilisation, improves performance,or offers resilience).

Mobile cloud computing would produce by 2016 45 billion dollars in revenue, according to the latest Visiongain research. There are several commercial prospects and research needs for mobile cloud computing as a result of this strong corporate market trend.

Multimedia search, sensor data applications, crowd computing, natural language processing, image processing,and social networking are some of the applications for which mobile clouds are in demand.

## 2. Related Works

There have been a slew of papers released recently discussing various aspects of cloud computing for mobile devices. In [20], for example, L. Zhong, B. Wang, and H. Wei from the perspective of mobile internet discussed mobile cloud computing. Mobile cloud computing has a set of six primary qualities. As a starting point, they are as follows: 1) overcome the limitations of terminal hardware 2) provide easy access to data 3) utilise intelligent load distribution and optimization, 4) improve the efficiency of task processing 5, eliminate regional restrictions and finally, 6) break through terminal hardware limitations. A comparison of existing mobile cloud computing models is also presented in [34].

In [19], Han Qi and Abdullah Gani use an integrated vision of cloud computing and mobile computing to describe mobile cloud computing. Among the mobile cloud infrastructures and architectures they looked at were Hyrax Infrastructure, Cloudlet and AlfredO Architecture. It's worth noting that they also highlight MCC's shortcomings in three ways: Mobile device restrictions, communication quality, and application service division are all factors that must be considered. It also discusses several open research questions in task division, data delivery, and service provision.[19] observes early developments in mobile media services cloud, which deliver applications and MC to more mobile subscribers than simply smartphone users. In addition, they point up several obstacles in MCC from three perspectives: constraints of mobile devices, communication quality, and division of application services. In addition, the article discusses various unresolved challenges in data delivery, work division, and service provision that need to be addressed.

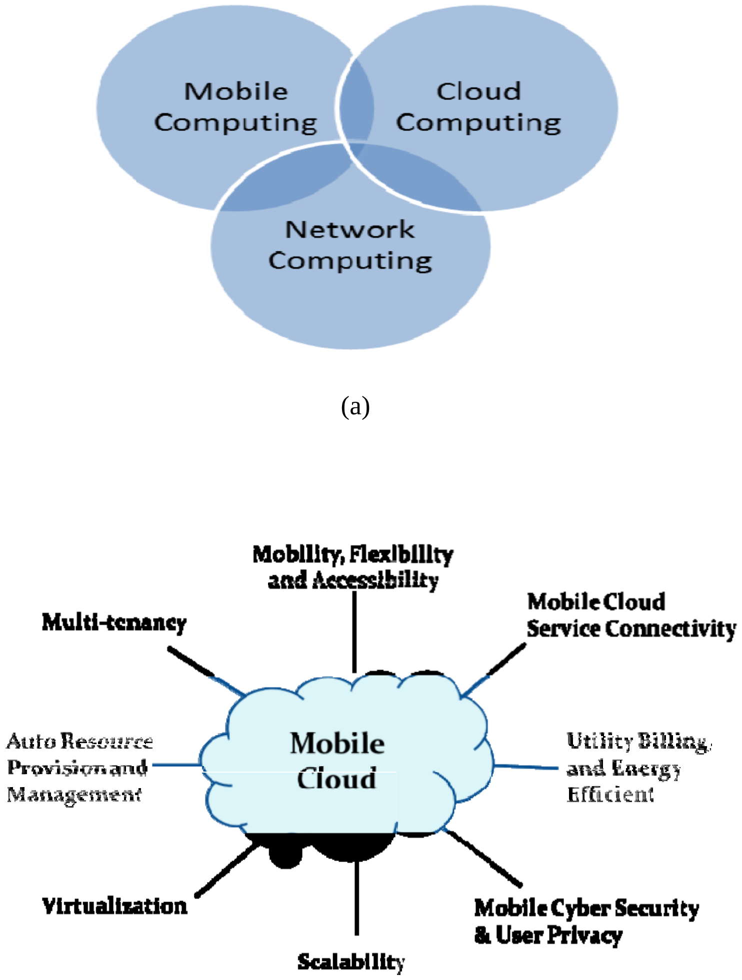
Cloud mobile media services are examined by S. Dey in [18] in terms of their early trends, growth, and benefits. In addition, he considers the effects and issues of privacy, energy consumption, cloud user experience, user and system reaction time, as well as cost and scalability. He investigates these issues and more in depth in his paper.In addition, the study discusses a number of relevant research avenues for cloud mobile media solutions.

Cloud computing may be integrated into the mobile environment using MCC, which solves performance issues is shown in a study by Hoang T. Dinh et al. [23]. (e.g. battery life, storage capacity, bandwidth) as well as environment, security (i.e. privacy, and reliability).

## 3. **Mobile Cloud Computing**

Many definitions of MCC are available, such as "a rich mobile computing system that uses unified elastic resources of diverse clouds and network technologies toward unconstrained functionality, storage, and mobility," for example. The pay-as-you-use principle [1] allows it to serve a large number of mobile devices across Ethernet or the Internet at any time and in any location.

According to the MCC forum, Mobile cloud computing, is simply an architecture in which data storage and processing take place externally to the mobile device. (http://www.mobilecloudcomputngforum.com/). By moving processing and to the cloud, data storage from mobile phones, MC applications as well as services are now available to many more mobile subscribers than just those with smartphones.




(b)
Figure 1: Concept of Mobile Computing

## 4. Results and Discussion

In the proposed research, we are using Mobile Cloud Computing dataset (and deploying it as a BigData file on cloud[Databricks]) and performing feature selection, extraction, correlation, classification and clustering of DataCenter/Virtual Machine based on the Mobile Cloud Data. And all this process is happening at DataBricks Cloud Environment.

### *4.1 Dataset used*

The dataset consists of the details of published journal papers. The dataset is uploaded as Big Data on Cloud. In this research DataBricks Cloud Environment is utilised. On the dataset feature selection and feature extraction are done. Count Vector Hashing is the method used for feature selection and feature extraction. Figure 1 gives the sample dataset. Figure 2 gives the Dataset after Count Vector Hashing Feature Selection/Extraction

| | Cites | Authors | Title | Year | Source | Publisher | ArticleURL | CitesURL | GSF |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 3.0 | A Abunaser, S Alshattnawi | Mobile cloud computing and other mobile techno... | 2012.0 | Journal of Mobile Multimedia | dl.acm.org | https://dl.acm.org/citation.cfm?id=2535625 | https://scholar.google.com/scholar?cites=17293... | |
| 1 | 1.0 | A Afianian, SS Nobakht... | Energy-efficient secure distributed storage in... | 2015.0 | ... (ICEE), 2015 23rd ... | researchgate.net | https://www.researchgate.net/profile/Amir_Afia... | https://scholar.google.com/scholar?cites=67727... | 3 |
| 2 | 3.0 | A Ahmad, A Paul, M Khan, S Jabbar... | Energy Efficient Hierarchical Resource Managem... | 2017.0 | IEEE Transactions ... | ieeexplore.ieee.org | http://ieeexplore.ieee.org/abstract/document/7... | https://scholar.google.com/scholar?cites=11507... | |
| 3 | 11.0 | A Ahmad, MM Hassan, A Aziz | A multi-token authorization strategy for secur... | 2014.0 | Mobile Cloud Computing ... | ieeexplore.ieee.org | http://ieeexplore.ieee.org/abstract/document/6... | https://scholar.google.com/scholar?cites=14743... | |
| 4 | 1.0 | A Ahmed, AA Hanan, K Omprakash... | Mobile Cloud Computing Energy-aware Task Offlo... | 2017.0 | ... : Proceedings of the ... | books.google.com | https://books.google.com/books?hl=en&lr=&id=EP... | https://scholar.google.com/scholar?cites=97324... | |
| 5 | 2.0 | A Ahmed, AS Sabyasachi, E Barlaskar | A survey on the QoS issue of Mobile Cloud Comp... | 2014.0 | 8th International Conference on ... | NaN | NaN | https://scholar.google.com/scholar?cites=31086... | 1 |
| 6 | 3.0 | A Ahmed-Nacer, MAN Samovar | Strong authentication for mobile cloud computing | 2016.0 | New Technologies for ... | ieeexplore.ieee.org | http://ieeexplore.ieee.org/abstract/document/7... | https://scholar.google.com/scholar?cites=70866... | |
| | | | QoS-based | | of Zululand | | | | |

Figure 2: Sample dataset

| | Cites | Year | GSRank | Volume | Issue | StartPage | EndPage | ECC | TT on Local SMD | TT in DCOF Based Computational Offloading | TT in Traditional Computational Offloading | Difference in TT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 3.0 | 2012.0 | 79.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1440.0 | 4876.0 | 2559.0 | 24331.0 | 89.482553 |
| 1 | 1.0 | 2015.0 | 347.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1089.0 | 5510.0 | 2902.0 | 28267.0 | 89.733612 |
| 2 | 3.0 | 2017.0 | 53.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1064.0 | 6566.0 | 3132.0 | 31609.0 | 90.091430 |
| 3 | 11.0 | 2014.0 | 60.0 | 0.0 | 0.0 | 0.0 | 0.0 | 445.0 | 6989.0 | 3345.0 | 35115.0 | 90.474156 |
| 4 | 1.0 | 2017.0 | 68.0 | 0.0 | 0.0 | 0.0 | 0.0 | 412.0 | 7406.0 | 3494.0 | 37010.0 | 90.559308 |
| 5 | 2.0 | 2014.0 | 177.0 | 0.0 | 0.0 | 0.0 | 0.0 | 336.0 | 7450.0 | 3757.0 | 38571.0 | 90.259521 |
| 6 | 3.0 | 2016.0 | 71.0 | 0.0 | 0.0 | 0.0 | 0.0 | 324.0 | 10414.0 | 3888.0 | 42244.0 | 90.796326 |
| 7 | 1.0 | 2017.0 | 63.0 | 0.0 | 0.0 | 0.0 | 0.0 | 305.0 | 11457.0 | 4379.0 | 47714.0 | 90.822400 |
| 8 | 1.0 | 2016.0 | 149.0 | 0.0 | 0.0 | 0.0 | 0.0 | 255.0 | 11857.0 | 4579.0 | 49481.0 | 90.745943 |
| 9 | 0.0 | 2015.0 | 258.0 | 0.0 | 0.0 | 0.0 | 0.0 | 255.0 | 13221.0 | 4864.0 | 54599.0 | 91.091412 |

Figure 3:Dataset after Count Vector Hashing Feature Selection/Extraction

7 out of 11 features are turn out to be constant means the remaining 4 feature have no impact on out main output or the variance of those 4 features are not good enough to count as constant. After Correlation analysis 1 of the 7 feature turn out to be lowest corelated at all which will be dropped. Figure 4 represents the Correlation Matrix of Remaining Data. Figure 5 represents the Current Working Dataset after all the Feature Selection/Extraction and Correlation check. Figure 6 represents the Dataset Description [mean/standard deviation/count/min /max/median etc.]

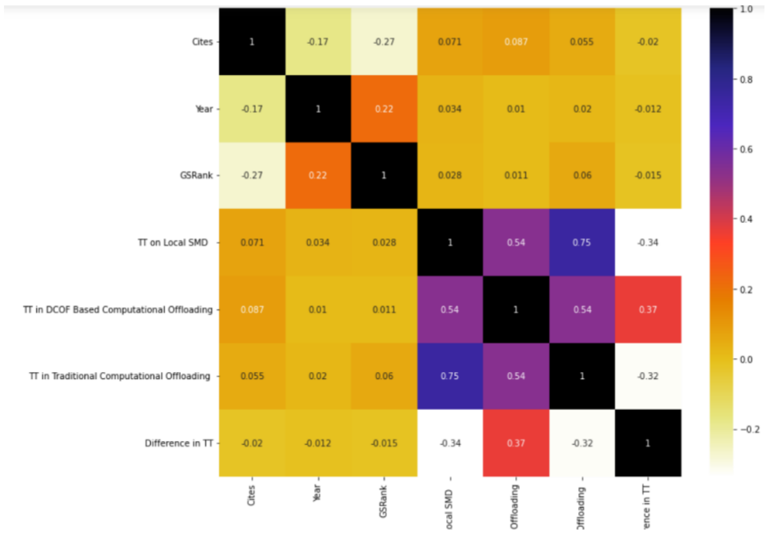


Figure 4: Correlation Matrix of Remaining Data

| | Cites | Year | GSRank | TT on Local SMD | TT in DCOF Based Computational Offloading | Difference in TT |
|---|---|---|---|---|---|---|
| **1790** | 2.0 | 2016.0 | 176.0 | 28000.0 | 1914.4 | 1433.400000 |
| **796** | 6.0 | 2015.0 | 4.0 | 4276.0 | 3889.0 | 78.743988 |
| **978** | 25.0 | 2012.0 | 27.0 | 51.0 | 68.0 | 99.052265 |
| **1544** | 1.0 | 2014.0 | 215.0 | 16950.0 | 6770.0 | 91.163610 |
| **1003** | 3.0 | 2014.0 | 143.0 | 22480.0 | 4380.0 | 96.540722 |
| **668** | 14.0 | 2013.0 | 124.0 | 63159.0 | 56467.0 | 69.320750 |
| **469** | 0.0 | 2017.0 | 143.0 | 13774.0 | 5222.0 | 91.142096 |
| **805** | 88.0 | 2013.0 | 17.0 | 13642.0 | 12177.0 | 74.003544 |
| **1002** | 0.0 | 2016.0 | 360.0 | 19182.0 | 3705.0 | 96.701066 |
| **160** | 0.0 | 2014.0 | 263.0 | 11457.0 | 4379.0 | 90.822400 |

Figure 5: Current Working Dataset after all the Feature Selection/Extraction and Correlation check

| | Cites | Year | GSRank | TT on Local SMD | TT in DCOF Based Computational Offloading | Difference in TT |
|---|---|---|---|---|---|---|
| **count** | 1238.000000 | 1238.000000 | 1238.000000 | 1238.000000 | 1238.000000 | 1238.000000 |
| **mean** | 11.152666 | 2014.666397 | 137.713247 | 17281.304891 | 11721.593624 | 4233.220407 |
| **std** | 46.216584 | 1.981161 | 96.848719 | 18604.916487 | 16328.532451 | 9284.839819 |
| **min** | 0.000000 | 2009.000000 | 1.000000 | 51.000000 | 4.600000 | 65.344865 |
| **25%** | 0.000000 | 2013.000000 | 56.000000 | 1767.000000 | 1572.400000 | 90.745943 |
| **50%** | 2.000000 | 2015.000000 | 119.500000 | 13221.000000 | 4898.000000 | 97.258401 |
| **75%** | 7.750000 | 2016.000000 | 211.000000 | 25687.000000 | 13182.000000 | 1601.840000 |
| **max** | 1218.000000 | 2018.000000 | 369.000000 | 99286.000000 | 91038.000000 | 43432.400000 |

Figure 6:: Dataset Description [mean/standard deviation/count/min /max/median etc.]

International Journal of Computer Science & Communication (ISSN: 0973-7391)
Volume 13 • Issue 1 pp. 37-45 Sept 2021 - March 2022 www.csjournals.com

*4.2 Machine Learning algorithms*

4.2.1 **K-NN Algorithm**

This technique is an instance-based learning method that classifies items based on the k training examples that are the closest to each other in the resource space.

**4.2.2 SVM Algorithm**

Supervised Learning techniques like as SVM (Support Vector Machine) are commonly used to solve classification and regression problems. As a classifier problem solver, it is primarily used in Machine Learning.

In the future, we will need to classify new data points more readily if we can locate the best line or decision boundary which can divide n-dimensional space into distinct classes.

Next we train our Model with KNN and Check Confusion Matrix

Table 1: Comparison of performance metrics of KNN and SVM algorithm

| Algorithm/ Performance metric | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| **KNN** | 0.82 | 0.78 | 0.62 | 0.91 |
| **SVM** | 0.65 | 0.13 | 0.48 | 0.21 |

KNN is Performing better in every aspect as compare to SVM so for our Regression and Clustering Process we will use KNN's Prediction value for calculation.

Performing Birch Clustering Method dividing Cloud Data(Cloud Hosts) into 7 Clusters(Datacenters) so we can assign which cloud host/virtual machine belongs to which datacenter. Figure 7 gives the Performance of BIRCH with and without global clustering.

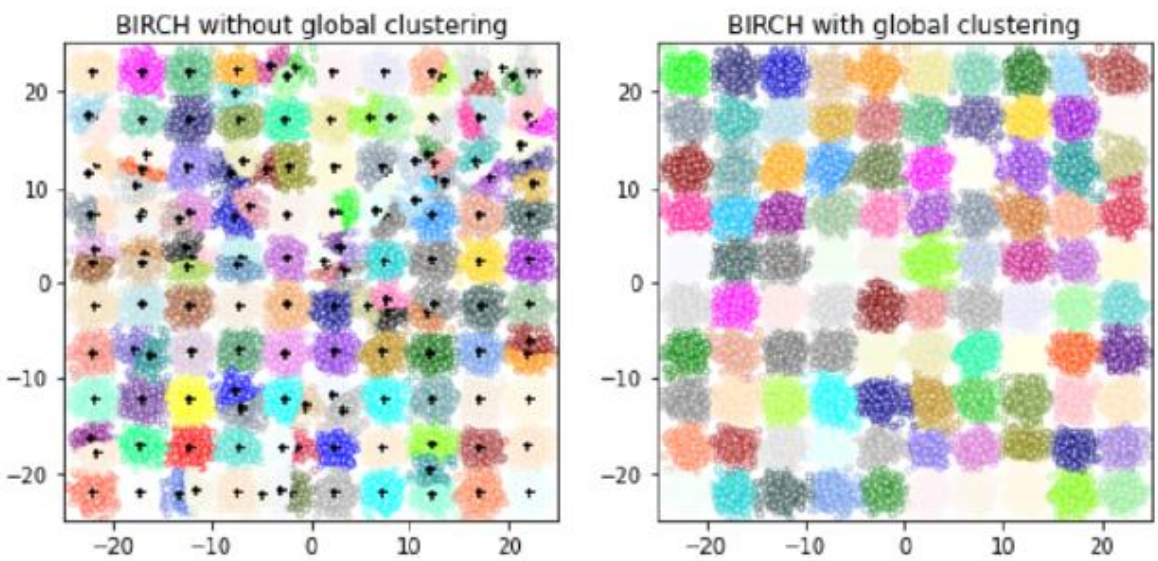



Figure 7:Performance of BIRCH with and without global clustering

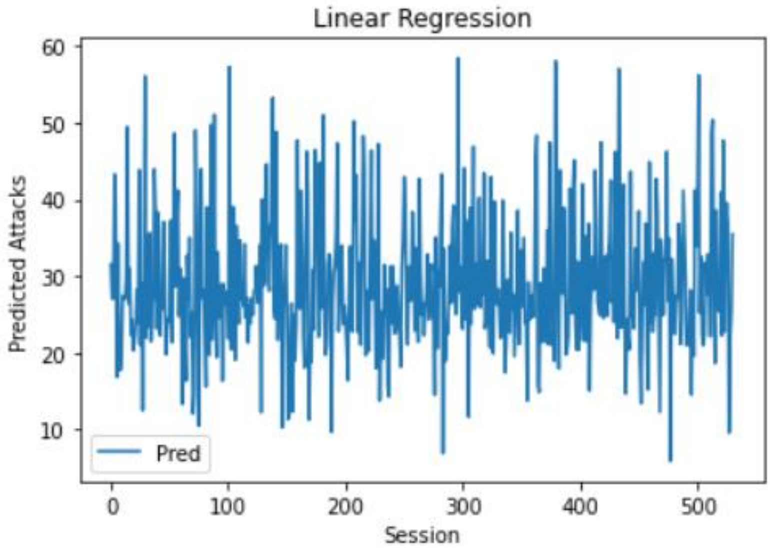



Figure 8: Predicted attacks using Linear regression

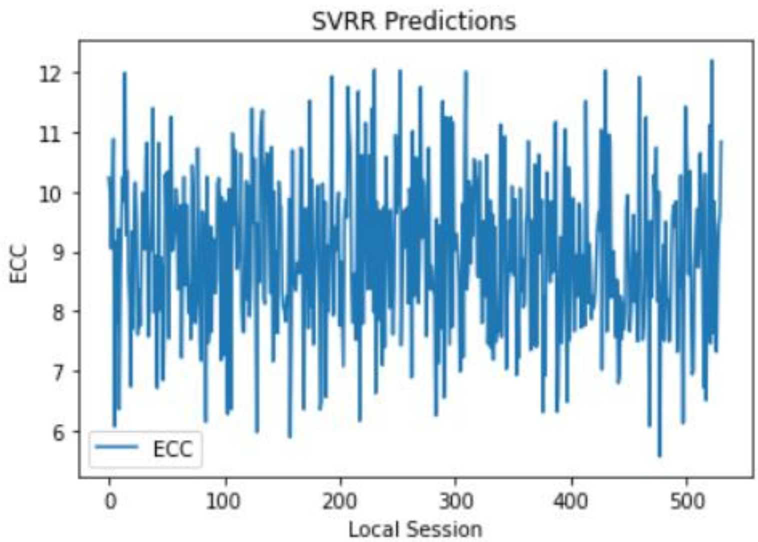



Figure 9: ECC for SVRR predictions

Figure 8 gives the predicted attacks using linear regression and Figure 9 gives the ECC for SVRR predictions after applying Linear Regression for SVM Predicted Data with our testing data.

*4.3 R2 Score Error*

The statistic displays the variance percentage of the dependent variable, which is used to determine the independent variable. On a handy 0–100 percent scale, it assesses the strength of the association between the dependent variable and regression model. The R2 score, also known as the co-efficient of determination. It always has a score between 0 and 100 percent. The response variable with a 0% score has no variability around its mean. The trained strong $R^2$score demonstrates its effectiveness. The latter compensates for the number of variables in a forecasting model and can be obtained as: If the new updated features are useful to the forecasting model, the rise in additional features can lead to an increase in $R^2$adjusted.

Table 2: R2 Score of SVM and KNN algorithms

| | KNN | SVM |
|---|---|---|
| R2 Score | -0.05 | -0.018 |

## 5. Conclusion

In the proposed research, we have used Mobile Cloud Computing dataset (and deploying it as a BigData file on cloud[Databricks]) and performing feature selection, extraction, correlation, classification and clustering of DataCenter/Virtual Machine based on the Mobile Cloud Data. And all this process is happening at DataBricks Cloud Environment. Machine learning algorithms, KNN and SVM are used for clustering and the better machine learning algorithm is considered. As we can see from the results, Liner Regression's Score is better then SVR that means Clustering using KNN is better as compare to SVM.

## REFRENCES

[1] Zohreh Sanaei, Saeid Abolfazli, Abdullah Gani, Muhammad Shiraz, "SAMI: Service-Based Arbitrated Multi-Tier Infrastructure for Mobile Cloud Computing", IEEE MobiCC'12 conference, MobiCC 2012:IEEE Workshop on Mobile Cloud Computing, Beijing, China

[2] Dijiang Huang, "Mobile Cloud Computing", IEEE COMSOC MMTC E-Letter, April, 2011.

[3] Z. Sanaei, S. Abolfazli, A. Gani, and R. H. Khokhar, "Tripod of requirements in horizontal heterogeneous mobile cloud computing," Proceedings of the 1st International Conference on Computing, Information Systems, and Communications, 2012.

[4] J. Kim, "Architectural patterns for service-based mobile applications," 2010 IEEE International Conference on Service-Oriented Computing and Applications (SOCA), pp. 1-4, 13-15 Dec. 2010.

[5] J. Mishra, S. K. Dash, and S. Dash, "Mobile-cloud: A framework of cloud computing for mobile application," Lecture Notes of the Institute for Computer Sciences, SocialInformatics and Telecommunications Engineering, vol. 86, pp. 347-356, 2012.

[6] P. Costa, M. Migliavacca, P. Pietzuch, and A. L. Wolf, "NaaS: Network-as-a-Service in the Cloud," Hot-ICE, 2012.

[7] "OpenStack Open Source Cloud Computing Software"; http://www.openstack.org/

[8] M. Mrissa, S. Tbahriti, and H. Truong, "Privacy Model and Annotation for DaaS," 2010 IEEE 8th European Conference on Web Services (ECOWS), pp. 3-10, 1-3 Dec. 2010.

[9] Mazedur Rahman, Jerry Gao, and Wei-Tek Tsai,"Energy Saving in Mobile Cloud Computing", Proceedings of IEEE International Conference on Cloud Engineering, San Francisco Bay, USA, 2013.

[10] Wei-Tek Tsai and Guanqiu Qi, "Mobile SaaS Framework", IEEE International Conference on Cloud Engineering (IC2E2013), San Francisco Bay, USA, March, 2013.

[11] Ashok Chandrasekar, Karthik Chandrasekar, Harini Ramasatagopan, Rafica Abdul Rahim, "Energy Conservative Mobile Cloud Infrastructure", Advances in Grid and Pervasive Computing, Lecture Notes in Computer ScienceVolume 7296, 2012, pp 152-161

[12] Lee Cocking, "The Future of Mobile Cloud Infrastructure", http://www.guardtime.com/2012/08/13/the-future-of-mobilecloud-infrastructure/

[13] Xiaodan Wei , Jian Zhang, Siduo Zeng, "Study of the Potential SaaS Platform Provider in China", 2009 WRI World Congress on Software Engineering, May, 2009.

[12] Dijiang Huang, Zhibin Zhou, Le Xu, Tianyi Xing, Yunji Zhong, "Secure Data Processing Framework for Mobile Cloud Computing",

[13] Dejan Kovachev Domink Renzel, Ralf Klamma, Yiwei Cao, "Mobile Community Cloud Computing: Emerges and Evolves", 2010 Eleventh International Conference on Mobile Data Management (MDM2010).

[14] Amreen Khan and KamalKant Ahirwar, "MOBILE CLOUD COMPUTING AS A FUTURE OF MOBILE MULTIMEDIA DATABASE", International Journal of Computer Science and Communication, Vol. 2, No. 1, January-June 2011.

[15] Mujtaba Hassanpur, Heindrick So, Jun Pan, "Survey of Mobile Cloud Computing Architectures & Infrastructures", Technical Report, San Jose State University, 2012.

[16] D. Koavchev, Y. Cao, R. Klamma, "Mobile Multimedia Cloud Computing and the Web," 2011 Workshop on Multimedia on the Web (MMWeb), pp. 21-26, 8-8 Sep. 2011.

[17] M. Mrissa, S. Tbahriti, and H. Truong, "Privacy Model and Annotation for DaaS," 2010 IEEE 8th European Conference on Web Services (ECOWS), pp. 3-10, 1-3 Dec. 2010.

[18] S. Dey, "Cloud Mobile Media: Opportunities, challenges, and directions,"2012 International Conference on Computing, Networking and Communications (ICNC), pp. 929-933, 30 Jan. 2012 - 2 Feb. 2012.

[19] Han Qi, Abdullah Gani, "Research on Mobile Cloud Computing: Review, Trend and Perspectives," Proceedings of the Second International Conference on Digital Information and Communication Technology and its Applications
(DICTAP), IEEE, pp. 195-202, Jun. 2012.

[20] L. Zhong, B. Wang, and H. Wei, "Cloud Computing Applied in the Mobile Internet," 2012 7th International Conference on Computer Science & Education (ICCSE), pp. 218-221, 14-17 July 2012.

[21] Niroshinie Fernando, Seng W. Loke, Wenny Rahayu, "Mobile cloud computing: A survey", Future Generation Computer Systems, Volume 29, Issue 1, January 2013, Pages 84–106.

[22] Hoang T. Dinh, Chonho Lee, Dusit Niyato, and Ping Wang, "A survey of mobile cloud computing: architecture, applications, and approaches", WIRELESS COMMUNICATIONS AND MOBILE COMPUTING, 2011.