USER-ANOMALY DETECTION IN MOBLIE NETWORK USING BIG DATA ANALYTICS

[1]*Vadipina Amarnadh,* [2] *R SaiSathwik Reddy,* [3] *P VivekVardhan,,*[4]*Bijarapu Shashank*

[1]*Assistant professor, Department of Computer Science and Engineering*
Anurag Group of Institutions, Hyderabad, Telangana, India

2,3,4Department of Computer Science and Engineering
Anurag Group of Institutions, Hyderabad, Telangana, India

Abstract-Having a mobile with a sustainable and safe mobile network is essential in this day and age. Modern-day companies and organisations use this data and perform predictions to provide a better experience to the user. We utilize the data from these mobile networks (Big data) that is, call detail records to find anomalous behaviour. We then compare these anomalies with ground truths to verify correctness. We also create an opportunity to build a prediction model to transform our anomalous data into anomaly-free data. This makes sure there are better datasets to perform better predictions are leading to a better user experience.

I. INTRODUCTION

With communication network evolving to 5G the data produced through the mobile network will be Big Data because of its Volume, Velocity and Variety.

Big Data analytical techniques are required to analyse and manipulate this Big Data generated from mobile networks. For small data, it was easier for us to identify and eradicate anomalies but with Big Data we require special techniques and methods. The need for the eradication of anomalies is high as it has become essential for organisations to provide top quality service and experience to users.

We identify all the anomalies in a mobile network from the call detail records using Hierarchical clustering and K-Means clustering.

II. PROPOSED WORK:

Here we propose the use of Hadoop and Map-Reduce with clustering approach. Around 10 million Call Detailed Records (CDRs) are used in this project, so processing of these huge amount of data takes longer time and not only the speed of processing is low but also the utilisation of the resources is not optimal. So we are using Hadoop, you can learn about Hadoop in article [4]. CDR(s) which we are using to find the user anomalies have noise, missing data etc, in order to remove these we applied Preprocessing[5]. After applying preprocessing the data which contains the missing data is replaced by the result of average applied on that entire column, this is the optimal way to fill the missing data values. In traditional technique a single processor was executing the entire task but in big data any and all tasks will be split and processed parallelly by these we achieved the proper utilisation of the resources. Here we used Map-reduce with k-means, map-reduce splits the huge CDR(s) into multiple chunks and distribute the chunks between map and reduce. Map will distribute these chunks to reducer and reduce will apply the logic of data that is to find the distance between each CDR and then generate the output and this output will be sent back to map, Map will aggregate all outputs and merge as a single total output. This output is stored in the distance.txt (text file). These data in the text file is used to apply k-means by

randomly choosing 10 centroids, will convert these CDR(s) to clusters. By seeing the clusters with CDR(s) we can conclude that the cluster with least number of call records is considered as the anomaly as many number of records are not anomalous and they will be grouped together within a cluster with large number of records. We also used Hierarchical clustering on the distance text file by applying agglomerative approach (bottom-up approach) with single-linkage strategy we obtained the clusters and compared both the clustering algorithm to get the better understanding of the clusters.

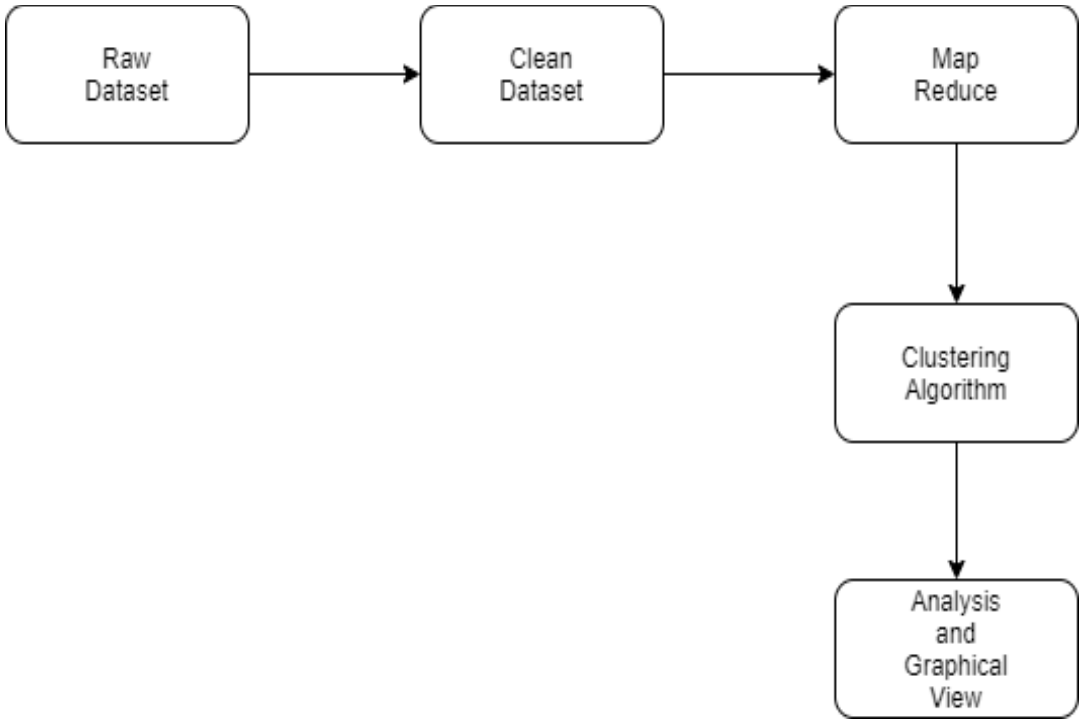

Figure 1. User anomaly detection using Big data Architecture Diagram

This is the architecture of our project. We took the raw dataset (CDRs) applied preprocessing to clean the dataset, then applied Map-Reduce with Clustering algorithm and Analysis and Graphical View of the anomalies. Wecreated a GUI using java libraries such as Swing and AWT for users to easily communicate with this project and can upload a dataset and apply preprocess, Generate centroids, K-means clustering, Hierarchical clustering, view the anomalies, dendrogram and chart in the GUI.

III. EXPERIMENT AND RESULT

The dataset taken for this experiment is call detail record (CDRs) of cellular network users. We apply clustering algorithms on data after preprocessing and selecting centroids. The resultant anomalous data records are shown in the form of table and also a chart

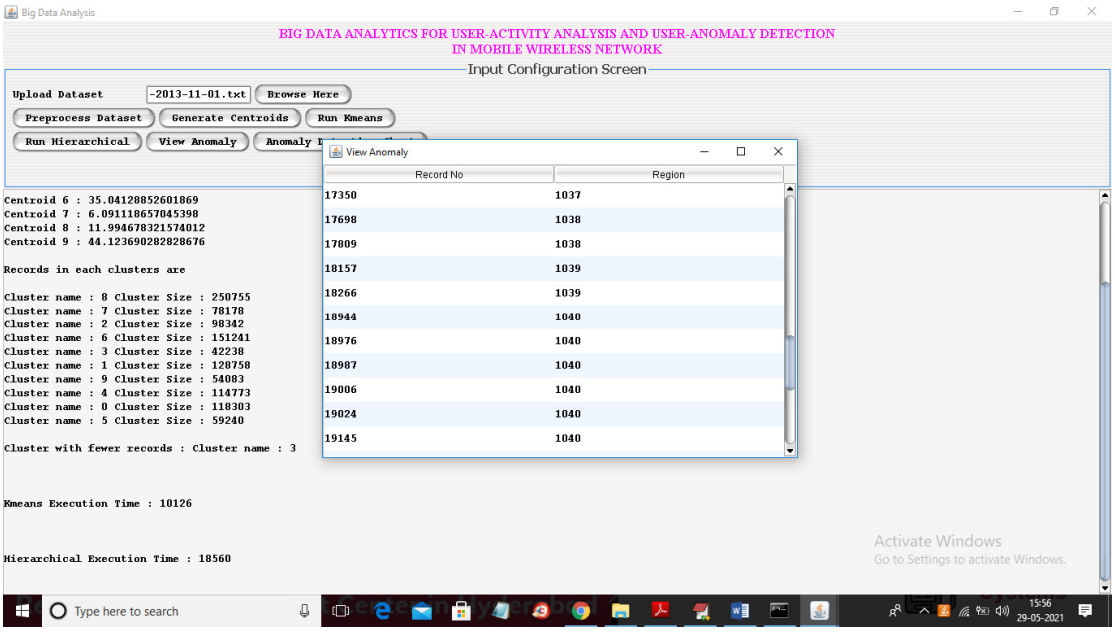

Figure 2. Anomalous data shown as a table

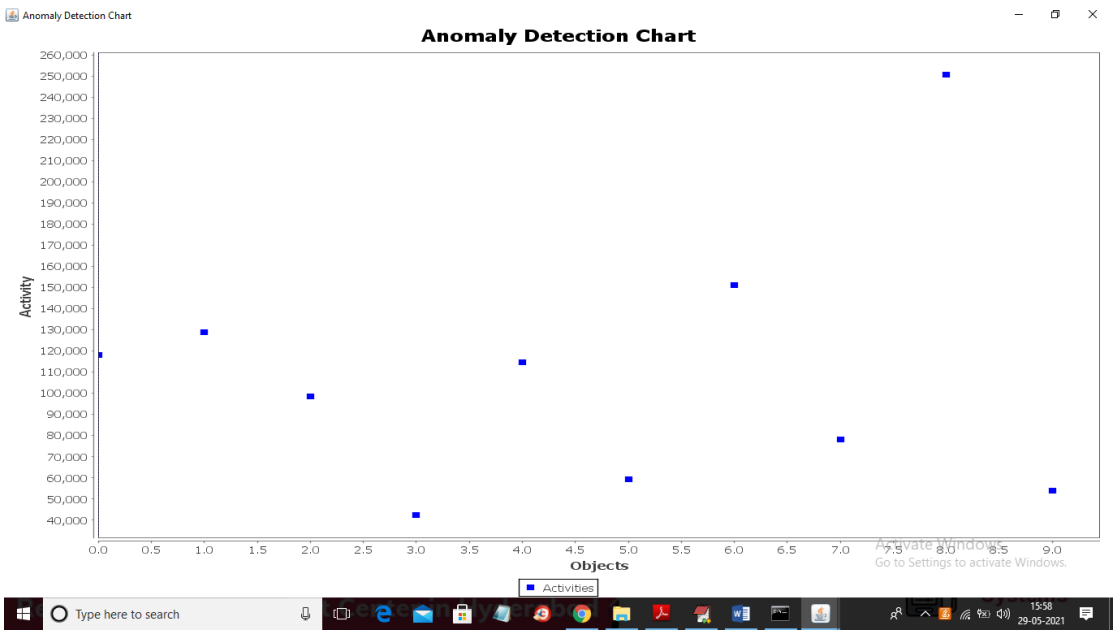

Figure 3. Anomalous data visualized as a chart

IV. CONCLUSION

In this paper we presented anomaly detection in mobile network big data (Call Detail Record: CDR) using machine learning (clustering) technique. We analyzed user activities at different time and location from the spatio-temporal information contained within CDR. Using k-means and hierarchical clustering techniques. The user activities that were unusually high caused unusual traffic demand and thus were categorized into anomalies. Thus with the help of anomaly detection, region of interests (RoI) can be identified, for which proper action

(e.g. proper resource allocation) can be taken in advance to meet the requirements. We also discussed the effect of anomalous and anomaly-free data by experimenting on a prediction model. We found that training the model with anomaly-free data resulted in less mean square error than with anomalous data.

REFERENCES:

1. D. Naboulsi, R. Stanica, and M. Fiore, “Classifying call profiles in largescale mobile traffic datasets,” in Proc. IEEE Conf. Comput. Commun., 2014, pp. 1806–1814.
2. A.M. Reddy, V. V. Krishna, L. Sumalatha and S. K. Niranjan, "Facial recognition based on straight angle fuzzy texture unit matrix," 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), Chirala, 2017, pp. 366-372, doi: 10.1109/ICBDACI.2017.8070865.
3. A.M. Reddy, K. SubbaReddy and V. V. Krishna, "Classification of child and adulthood using GLCM based on diagonal LBP," 2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), Davangere, 2015, pp. 857-861, doi: 10.1109/ICATCCT.2015.7457003.
4. Swarajya Lakshmi V Papineni, SnigdhaYarlagadda, HaritaAkkineni, A. Mallikarjuna Reddy. Big Data Analytics Applying the Fusion Approach of Multicriteria Decision Making with Deep Learning Algorithms *International Journal of Engineering Trends and Technology*, 69(1), 24-28, doi: 10.14445/22315381/IJETT-V69I1P204.
5. Srinivasa Reddy, K., Suneela, B., Inthiyaz, S.,Kumar, G.N.S., Mallikarjuna Reddy, A.” Texture filtration module under stabilization via random forest optimization methodology “International Journal of Advanced Trends in Computer Science and Engineering, Volume 8, No.3, May - June 2019
6. V. Soto and E. Fr´ıas-Mart´ınez, “Automated land use identification using cell-phone records,” in Proc. 3rd ACM Int. Workshop MobiArch. 2011, pp. 17–22.
7. [G. Munz, S. Li, and G. Carle, “Traffic anomaly detection using k-means ¨ clustering,” in Proc. GI/ITG Workshop MMBnet, 2007, pp. 1–8.
8. (available online) https://hadoop.apache.org/docs/current/
9. (available online) https://www.javatpoint.com/data-preprocessing-machine-learning
10. Mallikarjuna Reddy, V. Venkata Krishna, L. Sumalatha," Face Recognition based on Cross Diagonal Complete Motif Matrix", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.10, No.3, pp. 59-66, 2018.DOI: 10.5815/ijigsp.2018.03.07.
11. Ayaluri MR, K. SR, Konda SR, Chidirala SR. 2021. Efficient steganalysis using convolutional auto encoder network to ensure original image quality. PeerJ Computer Science 7:e356 https://doi.org/10.7717/peerj-cs.356