Available online at http://www.advancedscientificjournal.com
http://www.krishmapublication.com
IJMASRI, Vol. 2, issue 12, pp.799- 804, Dec. -2022
https://doi.org/10.53633/ijmasri

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY ADVANCED SCIENTIFIC RESEARCH AND INNOVATION (IJMASRI)

ISSN: 2582-9130 **IBI IMPACT FACTOR 1.5** **DOI: 10.53633/IJMASRI**

RESEARCH ARTICLE

DATA ANALYSIS USING ML ON GEOLOCATIONAL DATA

Vivek Chauhan[1] and Mrs. SeemaKolonia[2]

[1,2] Department of Information Technology, Maharaja Agrasen Institute of Technology affiliated to Guru Gobind Singh Indraprastha University, Rohini, Delhi

Abstract

Machine learning allows us to feed computer algorithms with large amounts of data and make computers analyze and make data-driven decisions and recommendations based solely on input data. This project will utilize ML to analyze geolocational data and user preferences to make smart recommendations to the user . In the fast-paced and busy environment that the average person lives in, it often happens that one is too tired to prepare a home-cooked meal. And of course, even if you get home cooked meals every day, it is not uncommon for you to want to have a good meal every now and then for social / recreational purposes. Now, imagine a scenario where someone has just moved to a new location. They already have certain preferences, certain tastes. It will save a lot of trouble for the student and food suppliers if the student lives near his favorite outlet. The convenience of the means better sales and time savings for customers. This project involves the utilization of K-Means Clustering to seek out the simplest accommodation for students in Bangalore (or the other city of your choice) by classifying

Keywords: ML, Unsupervised Learning, Simulation, Smart Recommendations, Data-driven approach

Introduction

In the fast-paced and busy environment where the average person lives, it is common for people to be too tired to prepare home-cooked meals. And of course, even if you eat homemade food every day, it's not unusual if you want to go out to eat once in a while for social/recreational purposes. However, it is a commonly understood idea that no matter where you live, the food you eat is an important aspect of the lifestyle you live. Now, imagine a scenario where someone has just moved to a new location. They already have certain preferences, certain tastes.

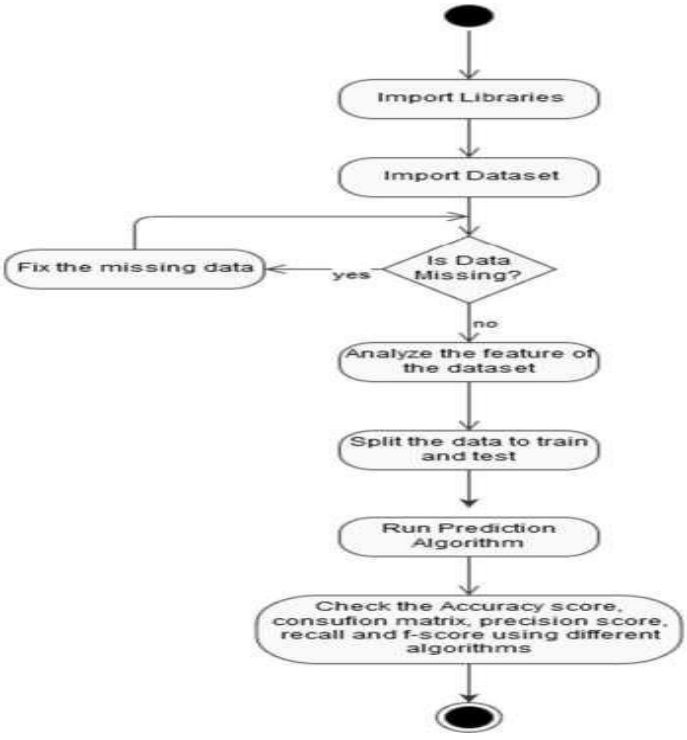
This will save students a lot of trouble and food suppliers if the student lives near his or her favorite outlet. This project involves using K-Means Clustering to find the best accommodation for students in Bangalore (or another city of your choice) by rating accommodation for incoming students based on their

preferences on facilities, budget and proximity to location.

Literature Review

There are many algorithms derived above to determine k automatically. Most of these methods are wrappers around kmeans or some other clustering algorithm for fixed k. Use the wrapper method divide and combine the rules for centers to increase or decrease the value of k as the algorithm progresses. After calculating the BIC or Bayesian Information Criterion(BIC is a method for scoring and selecting a model) for each clustering model. Apart from BIC, other scoring functions are also available. Some researchers use the MDL method to find the best k. The researchers also used the Minimum Description Length (MDL) framework, where the description length is the measurement value that tells us how well the data fit the model. This algorithm starts with a large value for k and removes the center (reduces k) each time that selection reduces the length of the description. Among the k reduction steps, they used the k means algorithm to optimize the fit of the model to the data.

Activity Diagram



Required tools:

- Python
- For Data – numpy and Pandas package
- For plotting – matpotlib package and seaborn packages
- For geospatial- geopy, folium
- For machine learning – sklearn (Preprocessing and cluster) scipy
- For deep learning

Feasibility Analysis:

The project Data Analysis using ML on Geolocational Data is a simple software application which is supported on a personal computer, just like any native application. It is a python based project with the renowned python libraries to manage the application. For an application as complex as an recommendation system using machine learning and AI , this software has simplistic approach and does not have many complex features therefore, it can be used at the bare-bones level. Sklearn (preprocessing and cluster) , scipy , matplotlib & seaborn packages, pandas packages for python are the other packages that we will be using in this application. Python provides us with many different features and services to complete our project. The convenient design of the geopy package for the plotting on map and the simple user experience will provide smooth and good experience tothe user.

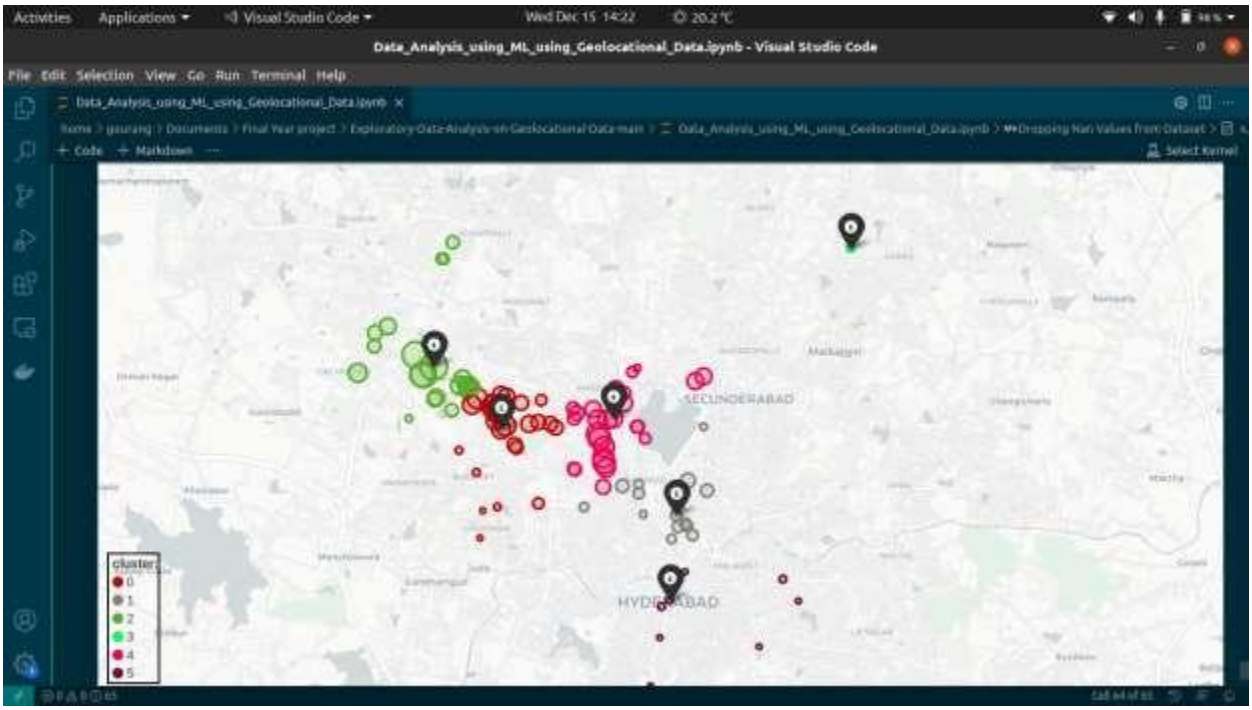
Complete work plan layout:

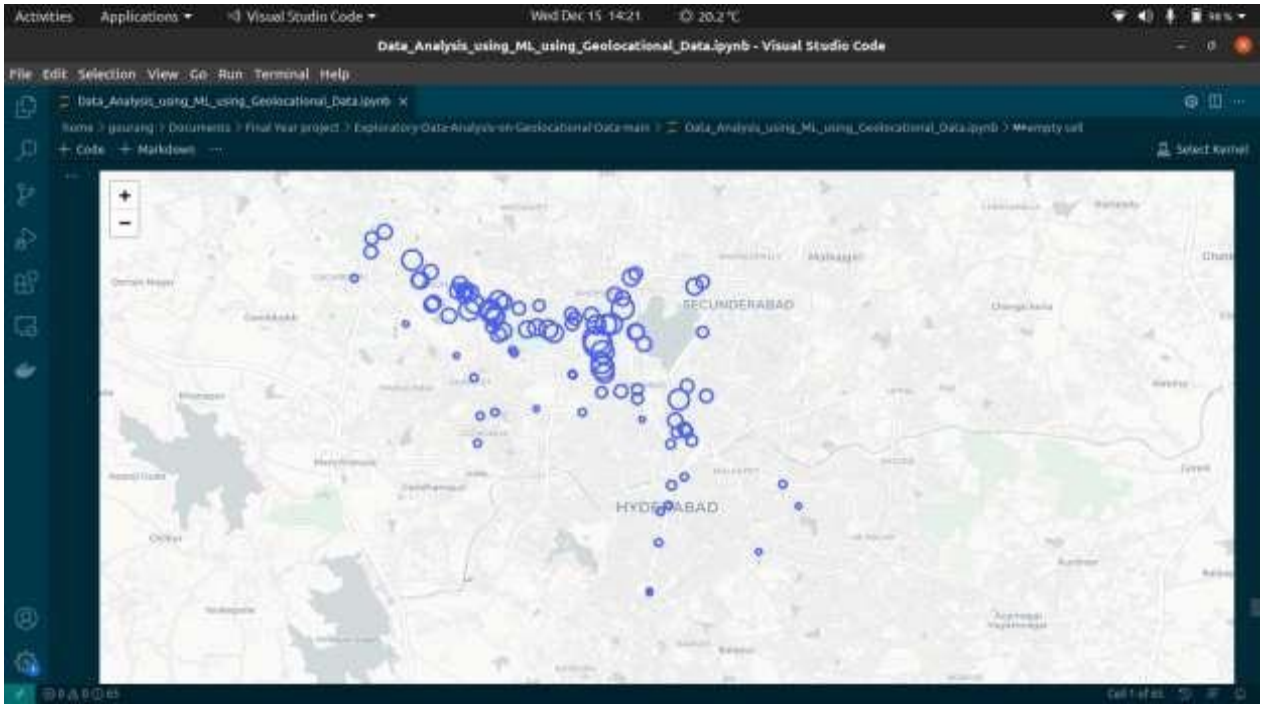
Modeling and Prediction with Machine Learning

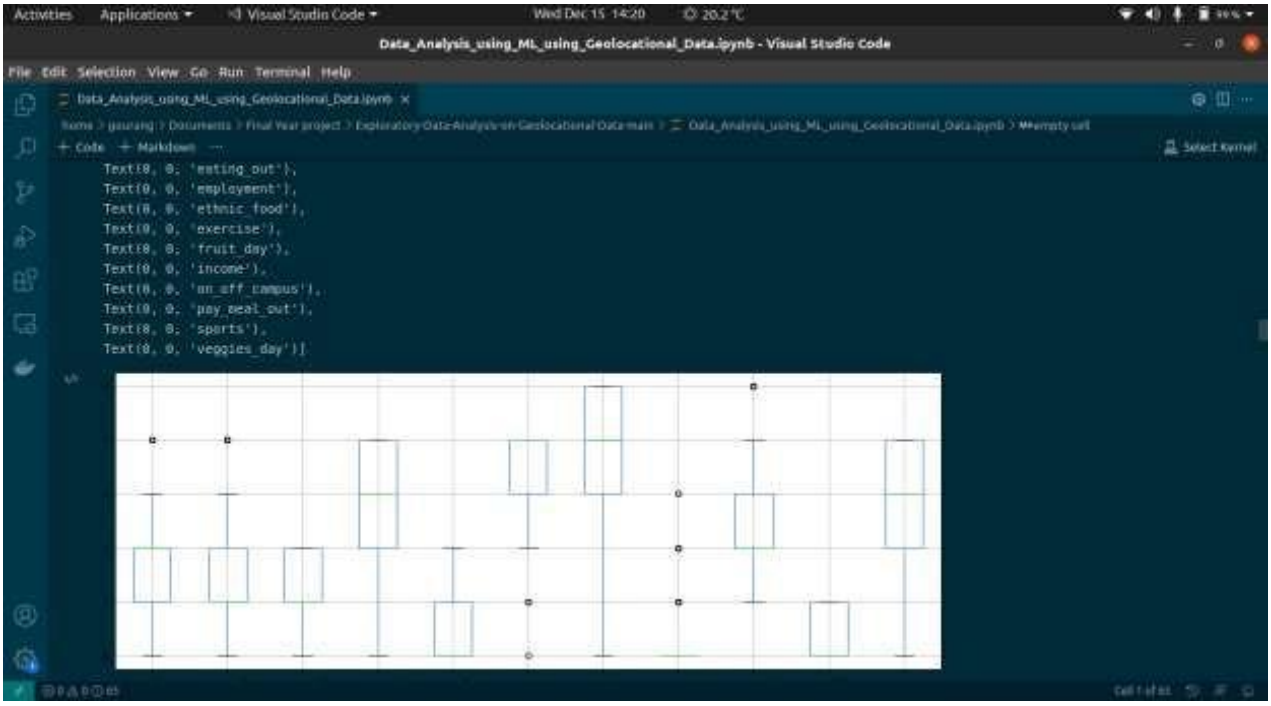
The main goal of the entire project is to predict the occurrence of heart disease with the highest accuracy. To achieve this, we will test several classification algorithms. This section covers all the results obtained from the study and introduces the best performer according to the accuracy metric. We have chosen a number of specific algorithms to

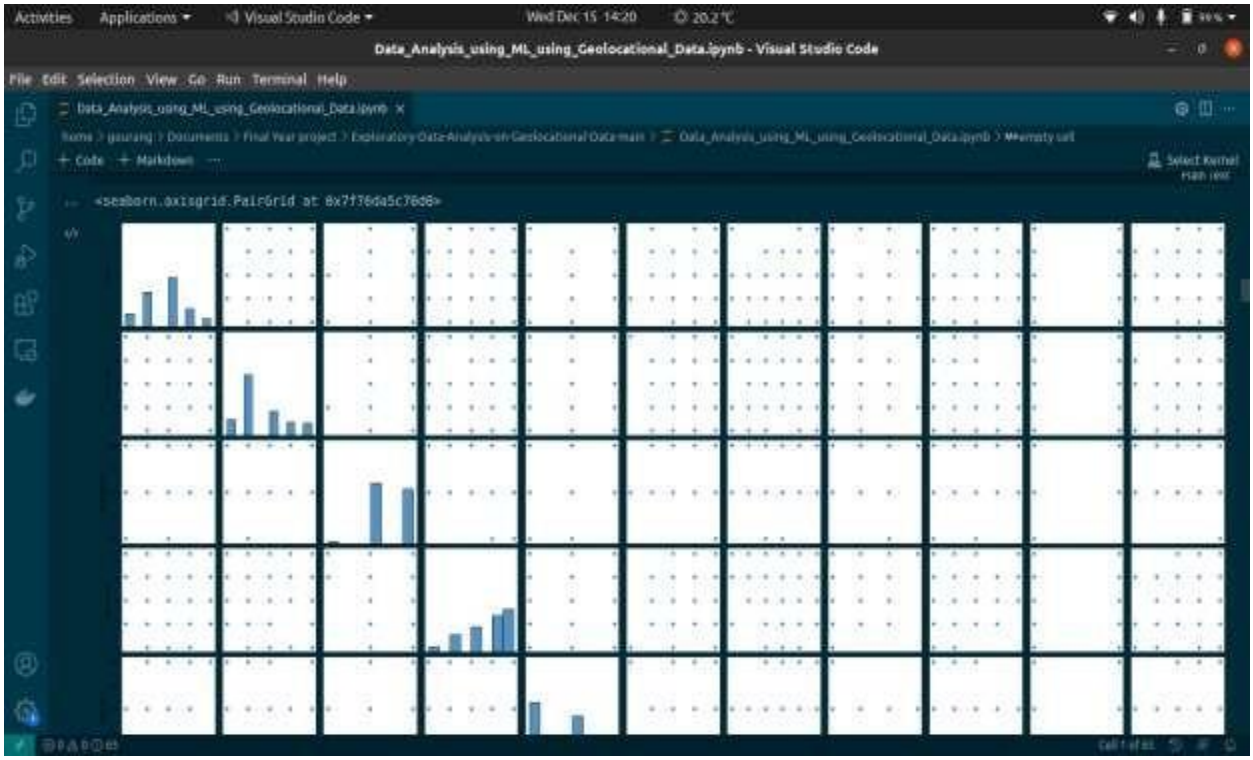
solve supervised learning problems across classification methods.

First, let's arm ourselves with a handy tool that benefits from the coherence of the SciKit Learn library and creates a common task for training our models. The reason we display accuracy on both the train and test sets is to allow us to evaluate whether the model over fits or under fits the data.









Reference

1. Saravanakumar Pichumani, T. V. P., Sundararajan, Rajesh Kumar Dhanaraj, Yunyoung Nam and Seifedine Kadry. (2021). "Ruzicka Indexed Regressive Homomorphic Ephemeral Key Benaloh Cryptography for Secure Data Aggregation in WSN," Journal of Internet Technology, vol. 22, no. 6 , pp. 1287-1297.
2. Kumar, D. R., Krishna, T. A and Wahi, A. (2018). Health Monitoring Framework for in Time Recognition of Pulmonary Embolism Using Internet of Things. Journal of Computational and Theoretical Nanoscience, 15(5), 1598–1602. https://doi.org/10.1166/jctn.2018.7347.
3. Krishnasamy, L., Dhanaraj, R. K. Ganesh Gopal, D. Reddy Gadekallu, T. Aboudaif, M. K and Abouel Nasr, E. (2020). A Heuristic Angular Clustering Framework for Secured Statistical Data Aggregation in Sensor Networks. Sensors, 20(17), 4937. https://doi.org/10.3390/s20174937.
4. Dhiviya, S., Malathy, S and Kumar, D. R. (2018). Internet of Things (IoT) Elements, Trends and Applications. Journal of Computational and Theoretical Nanoscience. 15(5), 1639–1643. https://doi.org/10.1166/jctn.2018.7354.
5. Rajesh Kumar, D and Shanmugam, A. (2017). A Hyper Heuristic Localization Based Cloned Node Detection Technique Using GSA Based Simulated Annealing in Sensor Networks. In Cognitive Computing for Big Data Systems Over IoT (pp. 307–335). Springer International Publishing. https://doi.org/10.1007/978-3-319-70688-7_13.
6. Prasanth, T., Gunasekaran, M and Kumar, D. R. (2018). Big data Applications on Health Care. 2018 4th International Conference on Computing Communication and Automation (ICCCA). 2018 4th International Conference on Computing Communication and Automation (ICCCA). https://doi.org/10.1109/ccaa.2018.8777586.
7. Rajesh Kumar, D and ManjupPriya S. (2013). Cloud based M- Healthcare emergency using SPOC. 2013 Fifth International Conference on Advanced Computing (ICoAC). 2013 Fifth International Conference on Advanced Computing (ICoAC). https://doi.org/10.1109/icoac.2013.6921965
8. Sathish, R and Kumar, D. R. (2013). Proficient algorithms for replication attack detection in Wireless Sensor Networks — A survey. 2013IEEE International Conference ON Emerging Trends in Computing, Communication and Nanotechnology (ICECCN). 2013 International Conference on Emerging Trends in Computing, Communication and Nanotechnology (ICE-CCN). https://doi.org/10.1109/ice-ccn.2013.6528465.
9. Soumya Ranjan Jena., Raju Shanmugam, Rajesh Kumar Dhanaraj and Kavita Saini. (2019). Recent Advances and Future Research Directions in Edge Cloud Framework. International Journal of Engineering and Advanced Technology, 9(2), 439– 444. https://doi.org/10.35940/ijeat.b3090.129219.
10. Kumar, R. N., Chandran, V. Valarmathi, R. S and Kumar, D. R. (2018). Bitstream Compression for High Speed Embedded Systems Using Separated Split Look Up Tables (LUTs). Journal of Computational and Theoretical Nanoscience, 15(5), 1719–1727. https://doi.org/10.1166/jctn.2018.7367.
11. Lalitha, K., Kumar, D. R., Poongodi, C and Arumugam, J. (2021). Healthcare Internet of Things – The Role of Communication Tools and Technologies. In Blockchain, Internet of Things, and Artificial Intelligence (pp. 331–348). Chapman and Hall/CRC. https://doi.org/10.1201/9780429352898-17.
12. Rajesh Kumar, D., Rajkumar, K. Lalitha, K and Dhanakoti, V. (2020).Bigdata in the Management of Diabetes Mellitus Treatment. In Studies in Big Data (pp. 293–324). Springer Singapore. https://doi.org/10.1007/978-981-15-4112-4_14.
13. Chandraprabha, M., and Dhanaraj, R. K. (2020). Machine learning based Pedantic Analysis of Predictive Algorithms in Crop Yield Management. 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA). 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA). https://doi.org/10.1109/iceca49313.2020.9297544

.
14. Dhanaraj, R. K., Rajkuma and Hariharan, U. (2020). Enterprise IoT Modeling: Supervised, Unsupervised, and Reinforcement Learning. In Business Intelligence for Enterprise Internet of Things (pp. 55–79). Springer International Publishing. https://doi.org/10.1007/978-3-030-44407-5_3.
15. Cynthia, J., Sankari, M. Suguna, M and Kumar, D. R. (2018).Survey on Disaster Management using VANET. 2018 4th International Conference on Computing Communication and Automation (ICCCA). 2018 4th International Conference on Computing Communication and Automation (ICCCA). https://doi.org/10.1109/ccaa.2018.8777331.
16. Dhanaraj, R. K., Shanmugam, A. Palanisamy, C and Natarajan, A. (2016). Optimal Clone Attack Detection Model using an Energy-Efficient GSA based Simulated Annealing in Wireless Sensor Networks. Asian Journal of Research in Social Sciences and Humanities, 6(11), 201. https://doi.org/10.5958/2249- 7315.2016.01186.2.
17. Sathya, K and Kumar, D. R. (2012, February). Energy efficient clustering in sensor networks using Cluster Manager. 2012 International Conference on Computing, Communication and Applications. 2012 International Conference on Computing, Communication and Applications (ICCCA). https://doi.org/10.1109/iccca.2012.6179177.
18. Lalitha, K., Varadhaganapathy, S., Santhoshi, S and Kumar, D. R. (2018, December). A Review on Possibilities of Hearing Loss and Implantable Hearing Devices for Teenagers. 2018 4th International Conference on Computing Communication and Automation (ICCCA). 2018 4th International Conference on Computing Communication and Automation (ICCCA). https://doi.org/10.1109/ccaa.2018.8777336.
19. Krishnamoorthi, S., Jayapaul, P and Dhanaraj, R.K. (2021). Design of pseudo- random number generator from turbulence padded chaotic map. Nonlinear Dyn (2021). https://doi.org/10.1007/s11071-021-06346-x.
20. Dhanaraj, R.K., Krishnasamy, L. Geman, O and Izdrui, D.R. (2021). "Black hole and sink hole attack detection in wireless body area networks," Computers, Materials & Continua, vol. 68, no.2, pp. 1949–1965, 2021. doi:10.32604/cmc.2021.015363.
21. Rajesh Kumar Dhanaraj., Lalitha, K. Anitha, S. Khaitan, S. Gupta, P and Goyal, M. K. (2021). Hybrid and dynamic clustering-based data aggregation and routing for wireless sensor networks [JB]. Journal of Intelligent & Fuzzy Systems,1–15.
22. Ramasamy, M.D., Periasamy, K. Krishnasamy, L. Dhanaraj, R.K. Kadry and, S and Nam, Y. (2020) "Multi-Disease Classification Model using Strassen's Half of Threshold (SHoT) Training Algorithm in Healthcare Sector," in IEEE Access, doi:10.1109/ACCESS.2021.3103746.
23. Ramakrishnan, V., Chenniappan, P. Dhanaraj, R. K. Hsu, C. H. Xiao, Y and Al-Turjman, F. (2021). Bootstrap aggregative mean shift clustering for big data anti-pattern detection analytics in 5G/6G communication networks. Computers &Electrical Engineering, 95, 107380.

*****