

A Comparative Analysis of Customer Behaviour Using Sophisticated Machine Learning

Dr. E. Nagarjuna¹, Dr. M. Suresh², L. Eswaramma³, P. Kiran⁴, M. Akshitha Laasya⁵

^{1, 2, 3, 4, 5}Department of Computer Science and Engineering

^{1, 2, 3, 4}QIS College of Engineering and Technology, Ongole, Andhra Pradesh, India

⁵CVR College of Engineering

e.nagarjuna@qiscet.edu.in¹, csehod@qiscet.edu.in², eswaramma.l@qiscet.edu.in³,

kiran.palakeeti@qiscet.edu.in⁴, 20B81A1206@cvr.ac.in⁵

Corresponding Author Mail: qispublications@qiscet.edu.in

Article Info

Page Number: 448 - 459

Publication Issue:

Vol 70 No. 2 (2021)

Article History

Article Received:

05 September 2021

Revised: 09 October 2021

Accepted: 22 November 2021

Publication: 26 December 2021

Abstract

Data is growing in leaps and bounds day by day with the rapid growth in cloud and internet technologies. Social media has become an integrated part of life as it is used to share content, opinions, reviews, suggestions etc. Anybody can voice out their opinions through social media. This data can be reused or analyzed to understand the customer and cater to his needs. In particular, e-Commerce platform enable the users to write product reviews to help other buyers purchase better products and be benefited from their opinions. The e-Commerce platform can also mine the data for sentiment analysis to understand the customer opinion and build recommendation systems to understand the customer needs and suggest products. Recommendation systems give a personalized shopping experience to users and there by increases sales and profits. In this project we use n-gram analysis to understand the data in the dataset and then apply multiple machine learning algorithms to the selected dataset to identify the appropriate classification algorithm that gives the best accuracy. We also try to build a recommender system that uses collaborative filtering to recommend appropriate products as per the customer needs.

Keywords: Data Visualization, Sentiment Analysis, Machine Learning, Natural Language Processing

I INTRODUCTION

Data is growing at a rapid rate due to social networking. Many users tend to voice their opinions on multiple topics and events in the form of reviews, blogs, tweets, posts etc. E-Commerce platforms enable their customers to post reviews about the performance and quality are various products that are being sold on the website. When customers give feedback about the products, it will help other customers make purchasing decisions. E-Commerce platforms and also take appropriate actions to maintain the quality of products that are being sold on their website. By analyzing the multiple reviews that are given to a product, the e-Commerce platform can find out the overall sentiment of the customers associated with that product. The e-Commerce platform can thus make a decision to keep a product are not based on the user sentiment. This will help to maintain the goodwill, quality of products which in turn improves the business of the e-Commerce platform. However, the reviews associated with the products might be thousand in numbers and it is not possible to

manually find the overall user sentiment as the system need to understand human language. Machine learning techniques and natural language processing techniques help in achieving this tedious task. In this project, we try to build a web application and a machine learning model that can accurately predict the sentiment from the review and a recommender system that can recommend suitable products to the customers.

Aim of the Project

We aim to extract the sentiment of the user for a particular product from the user review and get recommendations for users based on their purchases on the e-commerce platforms. The dataset used for this purpose is Amazon Reviews dataset and it is given to the developed machine learning model for training.

Scope of the Project

In this project, the accuracy of the developed machine learning model is calculated and compared with other machine learning algorithms and list of recommended products based on user processes is prepared. The admin of the application uploads and analyzes the data set and compare set with multiple machine learning algorithms. The process of maintaining users and their purchases on the e-Commerce platform does not fall under the purview of this project.

II LITERATURE SURVEY

“Consumer Behavior Modeling Based on Social Psychology and Complex Networks”

Bringing novelty in the services that an organization offers requires the clients to accept the changes. It is difficult to assess if a client would accept changes are not. This paper aims to bring about a new customer behavior framework which contains agents in a consumer network. This framework uses snowball sampling techniques. The remarkable of differentiating technologies that are used in this paper are social psychology and versatile networks.

“Modeling prediction in recommender systems using restricted Boltzmann machine”

Combined or collaborative filtering is a powerful technique that forms the core of recommender systems which in turn used in marketing of products to the internet users. This paper elaborates on controlled Boltzmann machine for combined filtering and suggests neighborhood and condition based model based on similarity and relevance scores. The evaluation of this model has been done based on hidden unit's rate of information and initializing functions. When a data set that has 22 million rows has been fed to the proposed model, it has achieved an accuracy of 78.5% in predicting user priorities in the recommended commercials.

III SYSTEM ANALYSIS

.Existing system:

Now a days, if an organization has to increase its profits, it has to understand the user intentions while trying to search or buy a product on the e-commerce websites. By understanding the user intentions and suggesting products, the organizations will be able to increase their sales and profits. However, in the existing scenario, the e-commerce websites are unable to accurately predict the

user requirements and unable to collate the user opinions about products based on reviews. We need to make use of sophisticated machine learning procedures like as Natural Language Processing (NLP)

Disadvantages:

- Not accurate
- Not capable of gathering consolidated user opinion or suggest products.

Proposed System:

In this project, we propose to develop an application that has the capabilities to suggest products based on consumer opinions and interests. We utilize the sophisticated machine learning techniques such as NLP to identify the sentiment of the user for a particular product based on the review. We also use collaborative filtering technique in our recommendation system to suggest products to users based on user purchases. These recommendations or suggestions can help them buy more products and thereby increase the sales.

Advantages:

- Automated recommendation system
- Improves sales and profits

IV IMPLEMENTATION

In order to implement the proposed application, we have divided it into 2 modules:

1. Admin
2. Consumer

Admin Module:

The admin of the application is accountable for performing the below actions:

1. Uploading the dataset
2. Statistical analysis of user opinions to gather the sentiment.
3. Development of machine learning model for sentiment analysis and recommender systems.
4. Evaluate the performance of various machine learning algorithms on the amazon products dataset.

Consumer Module:

The consumer of the system can utilize the machine learning services that are offered like:

1. Logging into the system
2. Write reviews for products and get the sentiment of the review.

3. Get product suggestions for purchase based on userId

IV SYSTEM DESIGN

Fig 1: Architecture diagram:

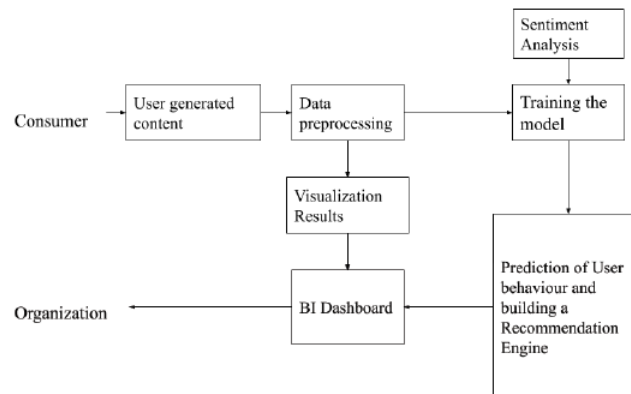


Fig 2 : SEQUENCE DIAGRAM:

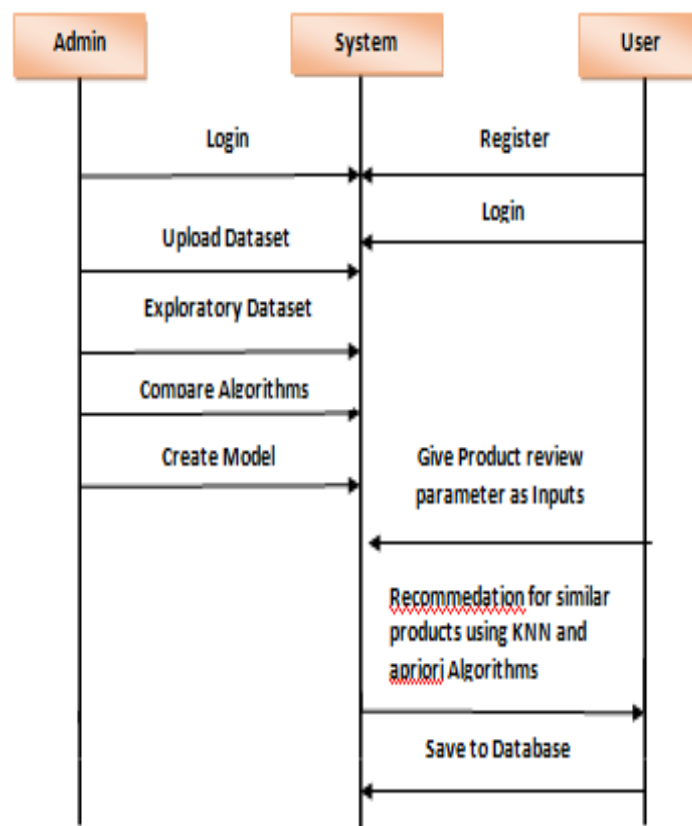


Fig 3: Data Flow Diagram: Admin

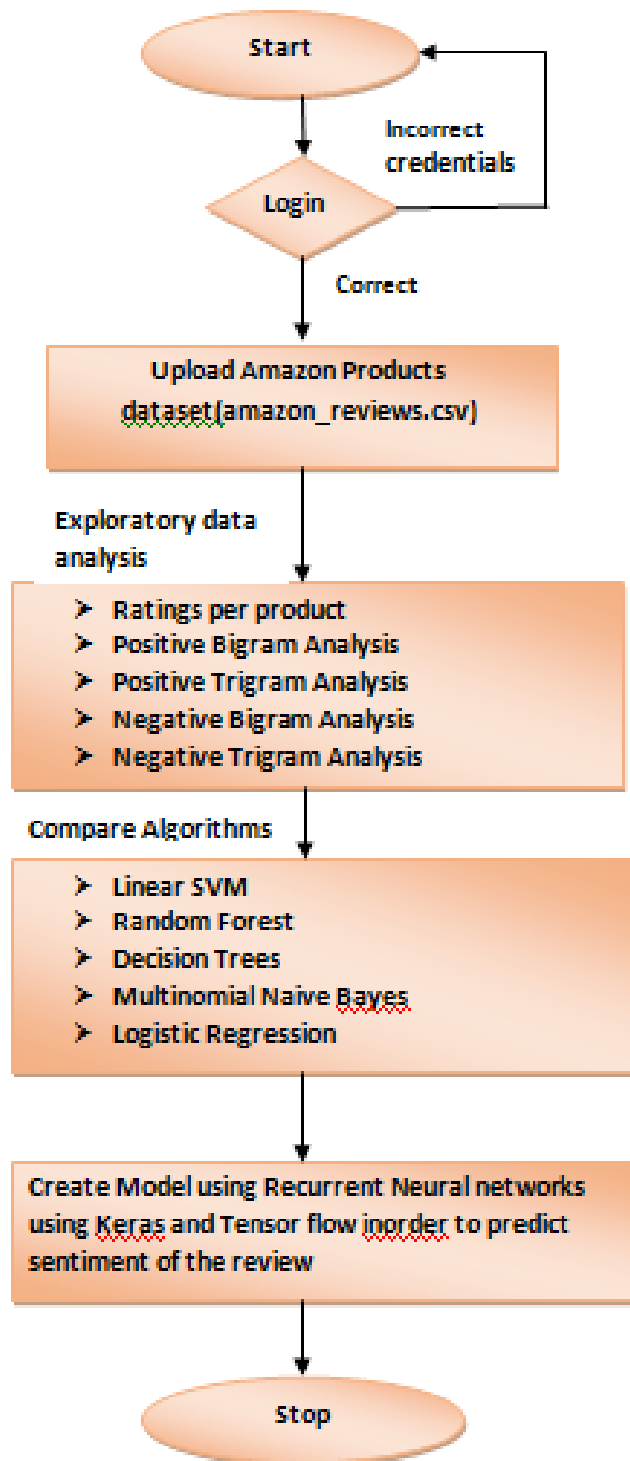
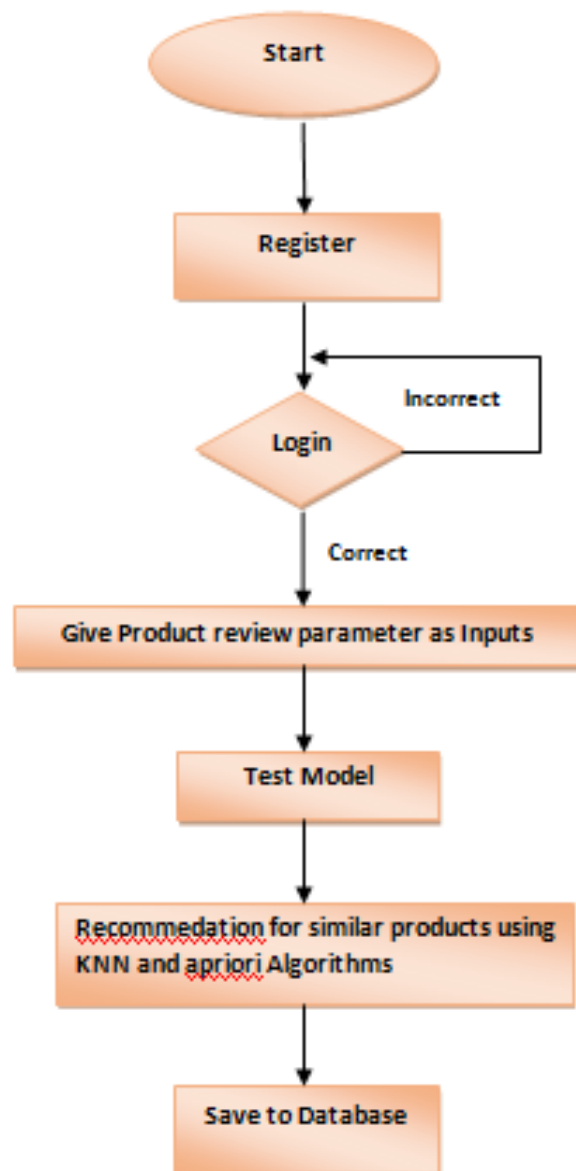


Fig 4: Data flow Diagram: User



V PROJECT EXECUTION

Fig 5: Negative Bigram:

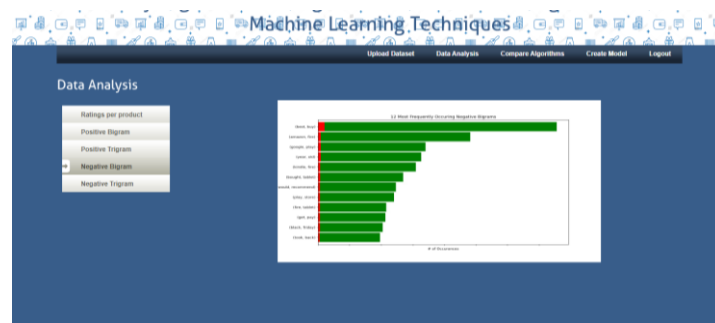


Fig 6: Negative Trigram:

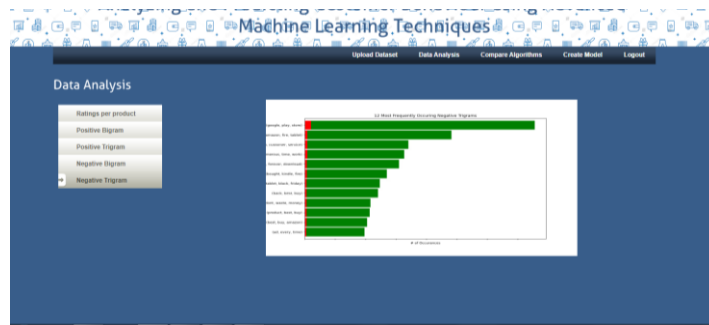


Fig 7: Compare Algorithms:

- **Logistic Regression:**

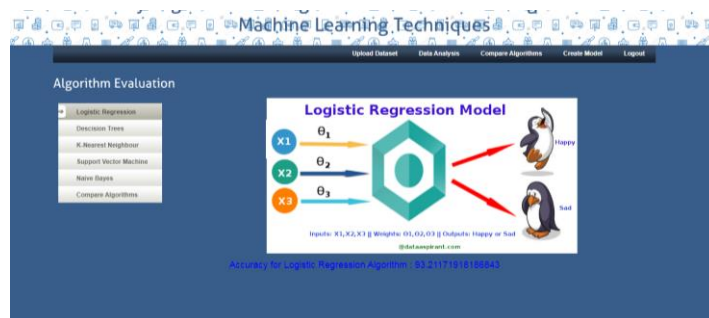


Fig 8: Decision Trees:

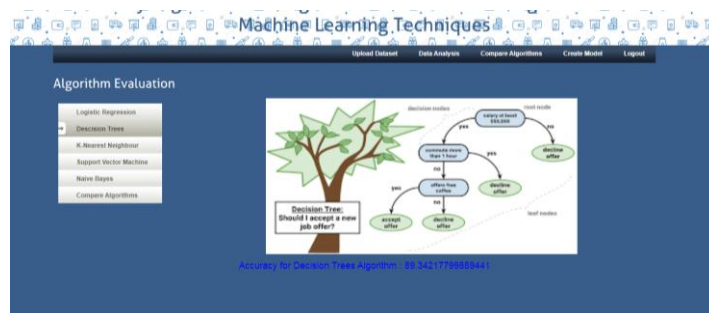


Fig 10: K-Nearest Neighbour:



Fig 12: Support Vector Machine:

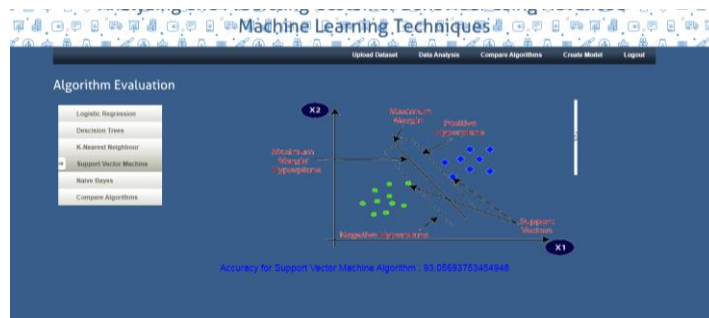


Fig 9: Naive Bayes:

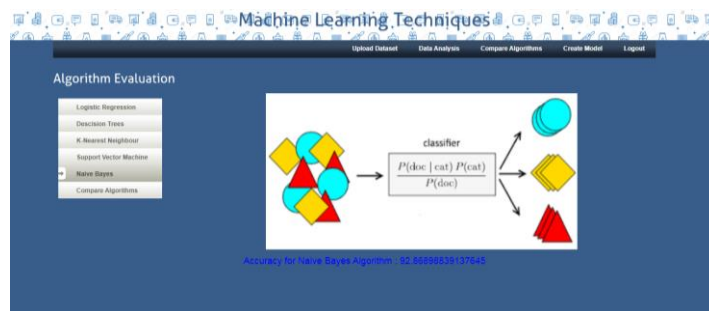


Fig 11: Evaluate Algorithms:

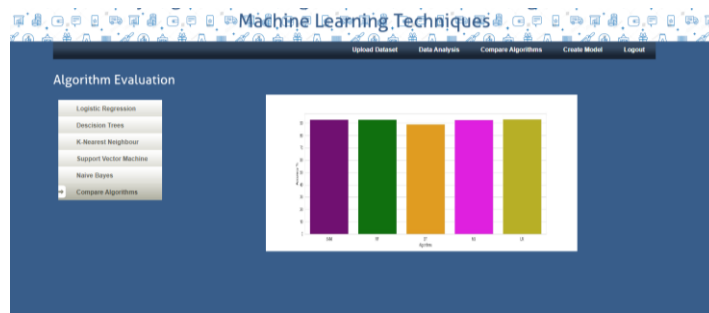


Fig 13: Sentiment Prediction

Machine Learning Techniques

Get Sentiment

Enter the review

Enter your review :

Submit

The Sentiment of
 the product that I have ever purchased
 is 82.0% positive !

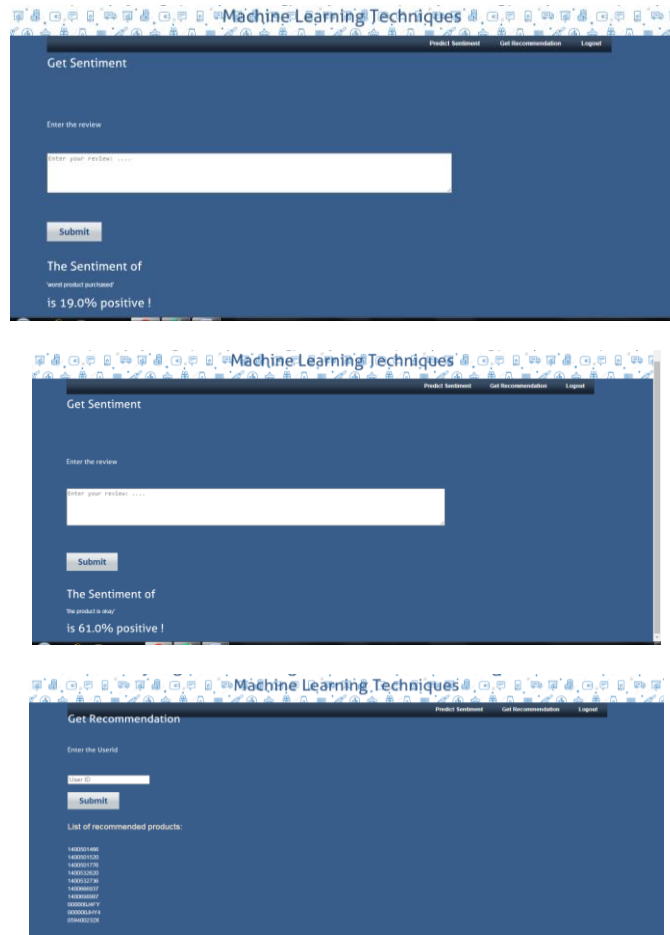


Fig 14: Product Recommendaion

VI CONCLUSION

The main objective of this project is to enable the eCommerce website to track customer experience, understand user opinion and suggest products based on consumer interest. This in turn would help to devise marketing strategies and business plans so as to improve their sales and profits. By analysing the reviews and search criteria the organisation can tap into the insights of customer opinions and interests. However, this would need sophisticated machine learning techniques such as NLP to gather customer opinion and suggest products based on their interests. In this project, we have used various machine learning techniques to get the sentiment of a product and set the data set to multiple machine learning algorithms to evaluate the accuracy. We observe that linear SVM give the best accuracy when compared to other algorithms and the recommender system employs kNN and prior algorithm. We also observed that the system performs better when compared to the existing systems. We have also deployed web application that enables the users to check the sentiment of a product based on product review and retrieve list of products that can be recommended to a particular user.

REFERENCES

- [1] T. Yoshida, M. Hasegawa, T. Gotoh, H. Iguchi, K. Sugioka and K. Ikeda, "Consumer behavior modeling based on social psychology and complex networks," The 9th IEEE International Conference on E-Commerce Technology and The 4th IEEE International Conference on

Enterprise Computing, E-Commerce and E-Services (CEC-EEE 2007), Tokyo, 2007, pp. 493-494. DOI: 10.1109/CEC-EEE.2007.36

- [2] R. He, J. McAuley. Modeling the visual evolution of fashion trends with one-class collaborative filtering. WWW, 2016
- [3] J. McAuley, C. Targets, J. Shi, A. van den Hengel. Image-based recommendations on styles and substitutes. SIGIR, 2015
- [4] Ben Yedder, Hanene & Zakia, Umme & Ahmed, Aly & Trajkovic, Ljiljana. (2017). Modeling prediction in recommender systems using restricted boltzmann machines. 2063-2068. 10.1109/SMC.2017.8122923.
- [5] Bengio, Y. & Delalleau, Olivier. (2008). Justifying and Generalizing Contrastive Divergence. Neural computation. 21. 1601-21. 10.1162/neco.2008.11-07-647.
- [6] Abdollahi, Behnoush & Nasraoui, Olfa. (2016). Explainable Restricted Boltzmann Machines for Collaborative Filtering.
- [7] Chamlerwat, & Bhatarakosol, & Rungkasiri, (2012). "Discovering consumer insight from twitter via sentiment analysis. journal of universal computer science." 18. 973-992.
- [8] S. K. Khatri and A. Srivastava, "Using sentimental analysis in prediction of stock market investment," 2016 5th International Conference on Reliability, Infocom Technologies, and Optimization (Trends and Future Directions) (ICRITO), Noida, 2016, pp. 566-569. DOI: 10.1109/ICRITO.2016.7785019
- [9] G. Xu, Y. Meng, X. Qiu, Z. Yu and X. Wu, "Sentiment analysis of comment texts based on BiLSTM," in IEEE Access, vol. 7, pp. 51522-51532, 2019. DOI: 10.1109/ACCESS.2019.2909919
- [10] Statnikov, Alexander I., Constantin F. Aliferis, Douglas P. Hardin, and Isabelle Guyon. "A gentle introduction to support vector machines in biomedicine: case studies." (2011).
- [11]. Agarwal, A., Harding, D.P. and Schumacher, J.R. (2004), "Organizing for CRM", McKinsey Quarterly, Vol. 3, pp. 80-91.
- [12] Aggelis, V. and Christodoulakis, D. (2005), "Customer clustering using RFM analysis", 9th WSEAS International Conference on Computers, Special Session Data Mining, Techniques and Application, Wisconsin.
- [13] Ali, O.G. and Arıtürk, U. (2014), "Dynamic churn prediction framework with more effective use of rare event data: the case of private banking", Expert Systems with Applications, Vol. 41 No. 17, pp. 7889-7903.
- [14] Bock, K.W.D. and Van, D.P.D. (2011), "An empirical evaluation of rotation-based ensemble classifiers for customer Churn prediction", Expert Systems with Applications, Vol. 38 No. 10, pp. 12293-12301.
- [15] Braun, M. and Schweidel, D.A. (2011), "Modeling customer lifetimes with multiple causes of churn", Marketing Science, Vol. 30 No. 5, pp. 881-902
- [16] Buckinx, W. and Van den Poel, D. (2005), "Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting", European Journal of Operational Research, Vol. 164 No. 1, pp. 252-268
- [17] Chan, C.C.H. (2008), "Intelligent value-based customer segmentation method for campaign management: a case study of automobile retailer", Elsevier Conference on Expert Systems with Applications, Vol. 34 No. 4, pp. 2754-2762.

- [18] Chen, S.H. (2016), "The gamma CUSUM chart method for online customer churn prediction", *Electronic Commerce Research and Applications*, Vol. 17 No. 1, pp. 99-111.
- [19] Chiang, D.A., Wang, Y.F., Lee, S.L. and Lin, C.J. (2003), "Goal-oriented sequential pattern for network banking churn analysis", *Expert Systems with Applications*, Vol. 25 No. 3, pp. 293-302.
- [20] Coussement, K. and De Bock, K.W. (2013), "Customer churn prediction in the online gambling industry: the beneficial effect of ensemble learning", *Journal of Business Research*, Vol. 66 No. 9, pp. 1629-1636.
- [21] Coussement, K. and Poel, D.V. (2009), "Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers", *Expert Systems with Applications*, Vol. 36 No. 1, pp. 6127-6134.
- [22] Devi Prasad, U. and Mahdi, S. (2012), "Prediction of churn behavior of bank customers using data mining tools", *Business Intelligence Journal*, Vol. 5 No. 1.
- [23] Dick, A.S. and Basu, k. (2003), "Customer loyalty: toward an integrated conceptual framework", *Journal of the Academy of Marketing Science*, Vol. 22 No. 2, pp. 99-113.
- [24] Freund, Y. and Schapire, R.E. (1996), "Experiments with a new boosting algorithm", 13th International Conference on Machine Learning, Murray Hill, NJ, pp. 148-156
- [25] Hadden, J., Tiwari, A., Roy, R. and Ruta, D. (2005), "Computer assisted customer churn management: state-of-the-art and future trends", *Computers and Operations Research*, Vol. 34 No. 10, pp. 2902-2917.
- [26] P Ramprakash, M Sakthivadivel, N Krishnaraj, J Ramprasath. "Host-based Intrusion Detection System using Sequence of System Calls" *International Journal of Engineering and Management Research*, Vandana Publications, Volume 4, Issue 2, 241-247, 2014
- [27] N Krishnaraj, S Smys. "A multihoming ACO-MDV routing for maximum power efficiency in an IoT environment" *Wireless Personal Communications* 109 (1), 243-256, 2019.
- [28] N Krishnaraj, R Bhuvanesh Kumar, D Rajeshwar, T Sanjay Kumar, Implementation of energy aware modified distance vector routing protocol for energy efficiency in wireless sensor networks, 2020 International Conference on Inventive Computation Technologies (ICICT), 201-204
- [29] Ibrahim, S. Jafar Ali, and M. Thangamani. "Enhanced singular value decomposition for prediction of drugs and diseases with hepatocellular carcinoma based on multi-source bat algorithm based random walk." *Measurement* 141 (2019): 176-183. <https://doi.org/10.1016/j.measurement.2019.02.056>
- [30] Ibrahim, Jafar Ali S., S. Rajasekar, Varsha, M. Karunakaran, K. Kasirajan, Kalyan NS Chakravarthy, V. Kumar, and K. J. Kaur. "Recent advances in performance and effect of Zn doping with ZnO thin film sensor in ammonia vapour sensing." *GLOBAL NEST JOURNAL* 23, no. 4 (2021): 526-531. <https://doi.org/10.30955/gnj.004020>, https://journal.gnest.org/publication/gnest_04020
- [31] N.S. Kalyan Chakravarthy, B. Karthikeyan, K. Alhaf Malik, D. Bujji Babbu, K. Nithya S. Jafar Ali Ibrahim, Survey of Cooperative Routing Algorithms in Wireless Sensor Networks, *Journal of Annals of the Romanian Society for Cell Biology*, 5316-5320, 2021
- [32] Rajmohan, G, Chinnappan, CV, John William, AD, Chandrakrishnan Balakrishnan, S, Anand Muthu, B, Manogaran, G. Revamping land coverage analysis using aerial satellite image mapping. *Trans Emerging Tel Tech.* 2021; 32:e3927. <https://doi.org/10.1002/ett.3927>

- [33] Vignesh, C.C., Sivaparthipan, C.B., Daniel, J.A. et al. Adjacent Node based Energetic Association Factor Routing Protocol in Wireless Sensor Networks. *Wireless Pers Commun* 119, 3255–3270 (2021). <https://doi.org/10.1007/s11277-021-08397-0>.
- [34] C Chandru Vignesh, S Karthik, Predicting the position of adjacent nodes with QoS in mobile ad hoc networks, *Journal of Multimedia Tools and Applications*, Springer US, Vol 79, 8445-8457, 2020