ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double,Blind Peer,Reviewed, Open,Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

Machine Learning Based Customer Churn Prediction

Virendrasingh Rajpurohit[1], Laukik Balai[2], Siddhi Nalavade[3], Prof. Mayuri Agrawal[4]

[1,2,3]*Student, Smt. Kashibai Navale College of Engineering, Pune, India*

[4]*Assistant Professor, Smt. Kashibai Navale College of Engineering, Pune, India*

Abstract: *Although substantial work has been devoted to improving the performance of customer churn prediction models, research on their interpretability and understanding of feature importance remains limited. However, for businesses to develop cost-effective customer retention strategies, it is vital to recognize which customer attributes are most influential in churn prediction. This paper explores the role of feature importance in forecasting churn within the telecommunications industry, utilizing a dataset containing nearly 40 customer attributes and employing three transparent, tree-based machine learning models. Specifically, we applied Random Forest, Decision Tree, and Extra Trees Classifier algorithms, leveraging their internal feature ranking mechanisms to evaluate the impact of individual features. We then compared the importance scores generated by these models to identify the most significant predictors overall. All three models achieved high classification performance, though the Decision Tree model demonstrated slightly lower accuracy than the others. Interestingly, the top-ranked features showed substantial overlap across the models, confirming their suitability for this analysis. Among the algorithms, Random Forest and Extra Trees Classifier, with their more intricate structures, provided superior results.*

Our findings highlight that variables such as customer tenure, service usage, and referral counts are the key determinants of churn, which aligns logically with customer engagement and loyalty patterns.

Keywords: Machine Learning, Feature Importance, Churn Prediction, Classification.

I. INTRODUCTION

Customer churn, defined as the discontinuation of a service or product by clients, presents a serious challenge for businesses across various industries. In the current competitive marketplace, focusing on customer retention is not only more cost-effective than acquiring new clients but also crucial for ensuring sustainable business growth and profitability. Accurately predicting customer churn enables companies to proactively implement tailored retention strategies, thus enhancing customer satisfaction and loyalty while minimizing financial losses.

The churn rate measures the percentage of customers who terminate their association with a business over a specified period. In simpler terms, it indicates the number of clients who either cancel their subscriptions or choose not to return for repeat business. A rising churn rate signifies a loss of customer base and potential revenue decline, whereas a falling churn rate reflects improved retention and customer loyalty, ultimately leading to business stability.

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double,Blind Peer,Reviewed, Open,Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

II. LITERATURE SURVEY

Arnesh Telukdarie and Mangaliso Maduna [1] proposed a structured framework consisting of four main stages: data preparation and processing, feature selection, dataset validation, and performance evaluationof predictive models.

B. Prabadevi and R. Shalini [2] analyzed historical customer data spanning approximately nine months before churn events. Their study reported classification accuracies of 83.9%, 82.6%, 82.9%, and 78.1% across different algorithms, highlighting the potential of machine learning in this domain.
Suhel Malik, Siddhart Runwal, and colleagues [3] presented a comprehensive review of churn prediction models, focusing on techniques such as logistic regression, decision trees, random forests, and neural networks, comparing their effectiveness in predicting customer attrition.

R. Srinivasan and D. Rajeswari [4] examined various machine learning methods for building churn models tailored for telecom operators, helping them identify customers likely to discontinue services. Their comparative analysis identified the most efficient algorithmic approaches.

Pulin Yang et al. [5] initiated their study by visualizing data before applying multiple classification models, including Random Forest, Support Vector Machines (SVM), and Gradient Boosted Decision Trees (GBDT). Their results demonstrated the superior performance of Random Forests in terms of classification accuracy.

Ke Peng, Yen Peng, and Wengaung Li [6] utilized Shapley value-based explanations to interpret model outputs and emphasized features such as the total number and amount of transactions over the past year as strong predictors of customer churn.

Peddarapu Rama Krishna and team [7] aimed to evaluate different models for predicting churn with high precision. Their work emphasized the importance of proactively identifying potential churners to improve customer retention, comparing various machine learning algorithms for optimal results.

Yana Fareniuk et al. [8] identified the main drivers behind customer churn and proposed that predictive models can help businesses minimize customer loss, optimize services, and improve overall business outcomes.

Oleksandr Dluhopolsky and collaborators [9] conducted a comparative study of multiple data science approaches for classifying customers by their likelihood of churn, helping organizations predict and manage customer attrition more effectively.

Jainam D and his research group [10] outlined the process of developing churn prediction systems, exploring the challenges and suggesting practical solutions to mitigate churn rates and boost customer retention.

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double,Blind Peer,Reviewed, Open,Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

III. SYSTEM LAYOUT

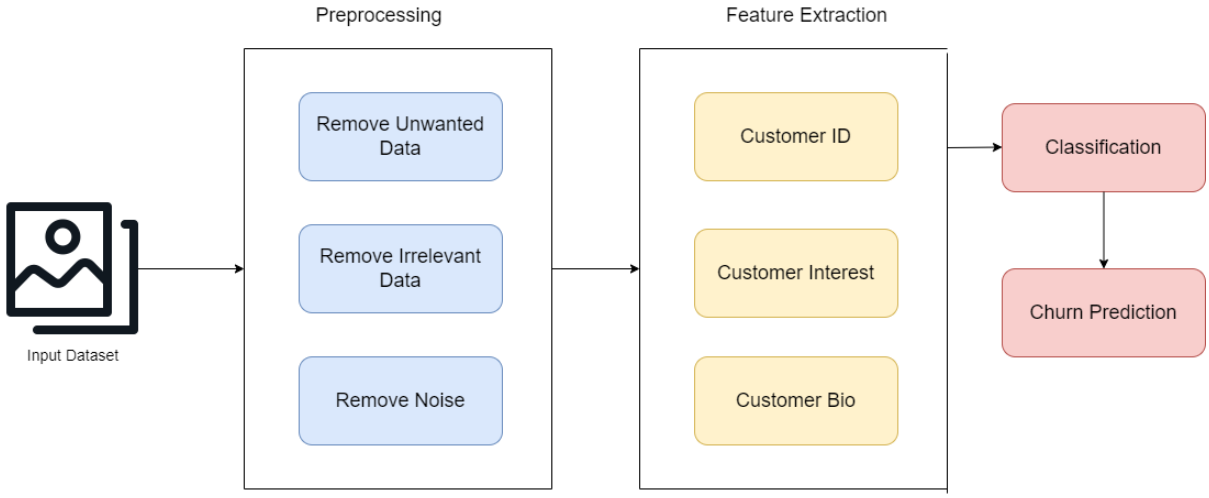

Figure 1: System Architecture

IV. ALGORITHM – RANDOM FOREST

The algorithm used here is Random Forest. Random Forest is the most popular and powerful algorithm of machine learning.

- Step 1: Assume N as number of training samples and M as number of variables within the classifier.
- Step 2: The number m as input variables to decide the decision at each node of the tree; m should be much less than M.
- Step 3: Consider training set by picking n times with replacement from all N available training samples. Use the remaining of the cases to estimate the error of the tree, by forecasting their classes.
- Step 4: Randomly select m variables for each node on which to base the choice at that node. Evaluate the best split based on these m variables in the training set.
- Step 5: Each tree is fully grown and not pruned (as may be done in constructing a normal tree classifier).

  For forecasting, a new sample is pushed down the tree. It is assigned the label of the training sample in the terminal node it ends up in. This procedure is repeated over all trees in the ensemble, and the average vote of all trees is reported as random forest prediction. i.e. classifier having most votes.

V. RESULTS AND DISCUSSION

Home Page:

The screenshot is of a home page. An user needs to signup first in order to find the churn result. Then user can login and fill out the questionnaire that appears on the screen and then the result will be dispalyed.

 DOI: 10.5281/zenodo.15617349

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double,Blind Peer,Reviewed, Open,Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

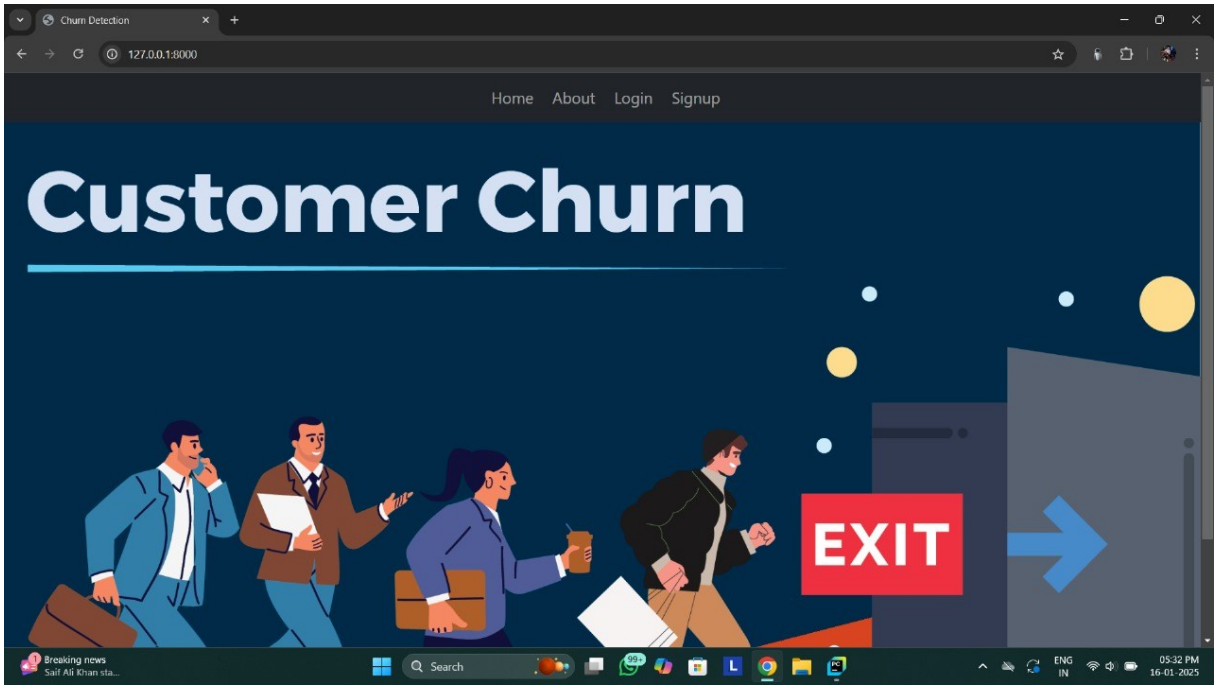

Figure 2: Home Page

Signup Page:

The screenshot displays the Signup page where an user can signup by entering fields like Username, First Name,Surname and Password.

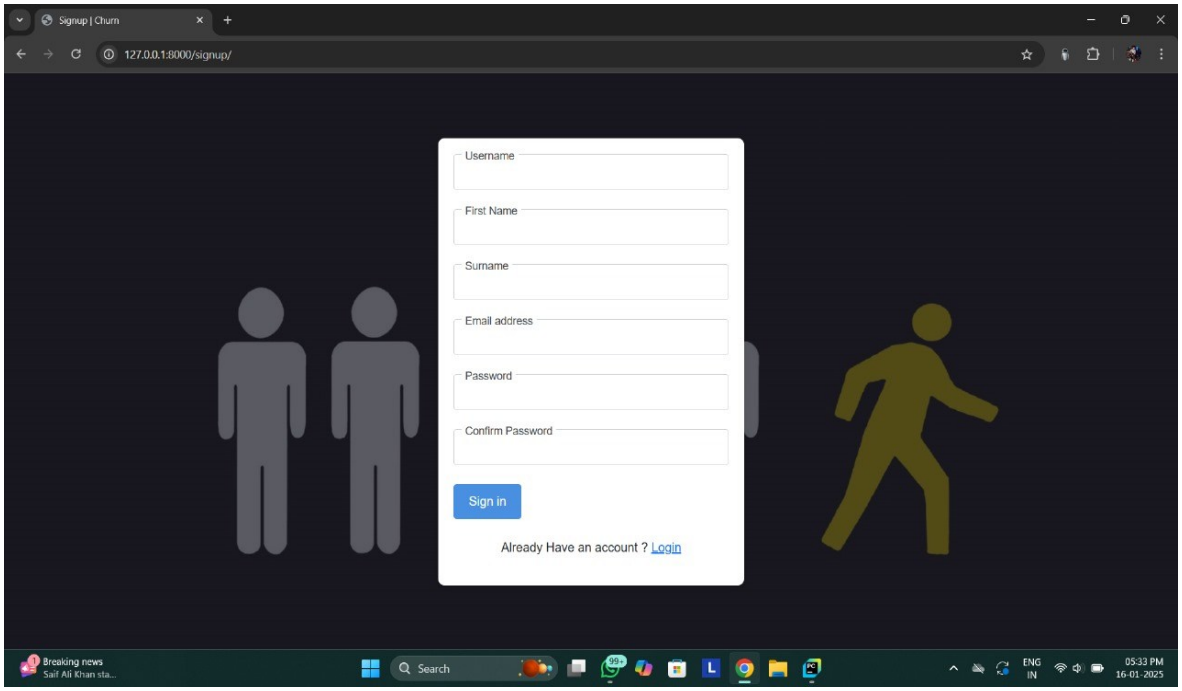

Figure 3: Signup Page

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double,Blind Peer,Reviewed, Open,Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

Login Page:

The screenshot displays the login page using which an user who is already a member can login using username/email and password.

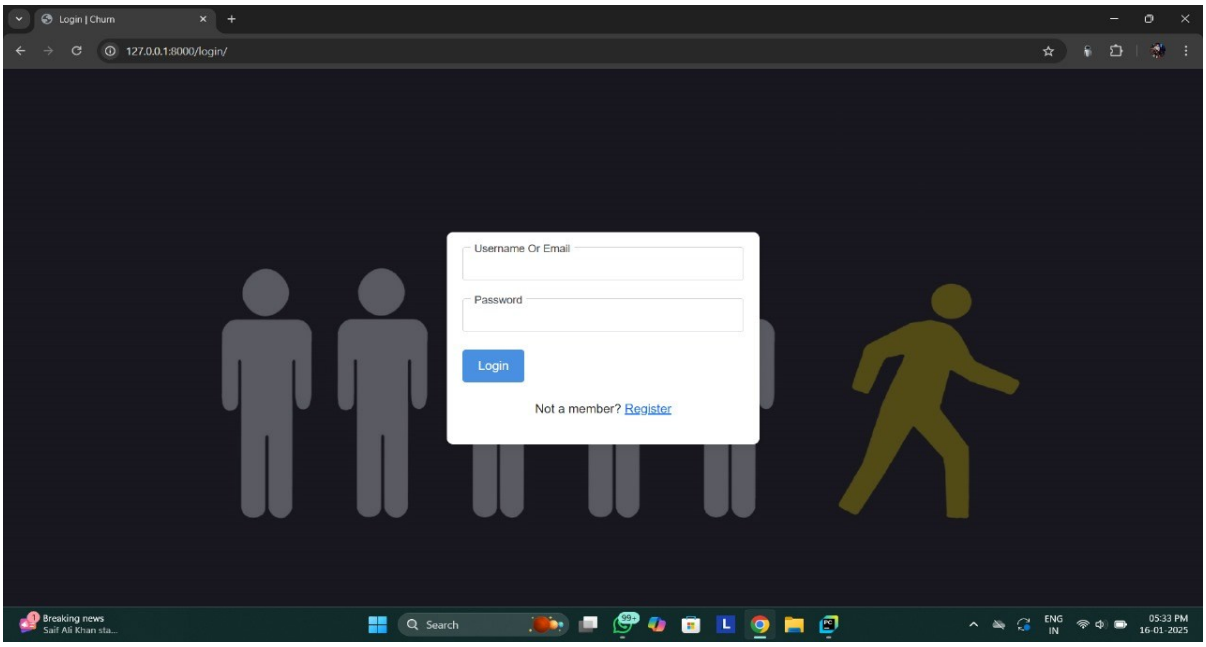

Figure 4: Login Page

The sreenshot 8.4 appears after login. It displays the username entered by the user during signup.

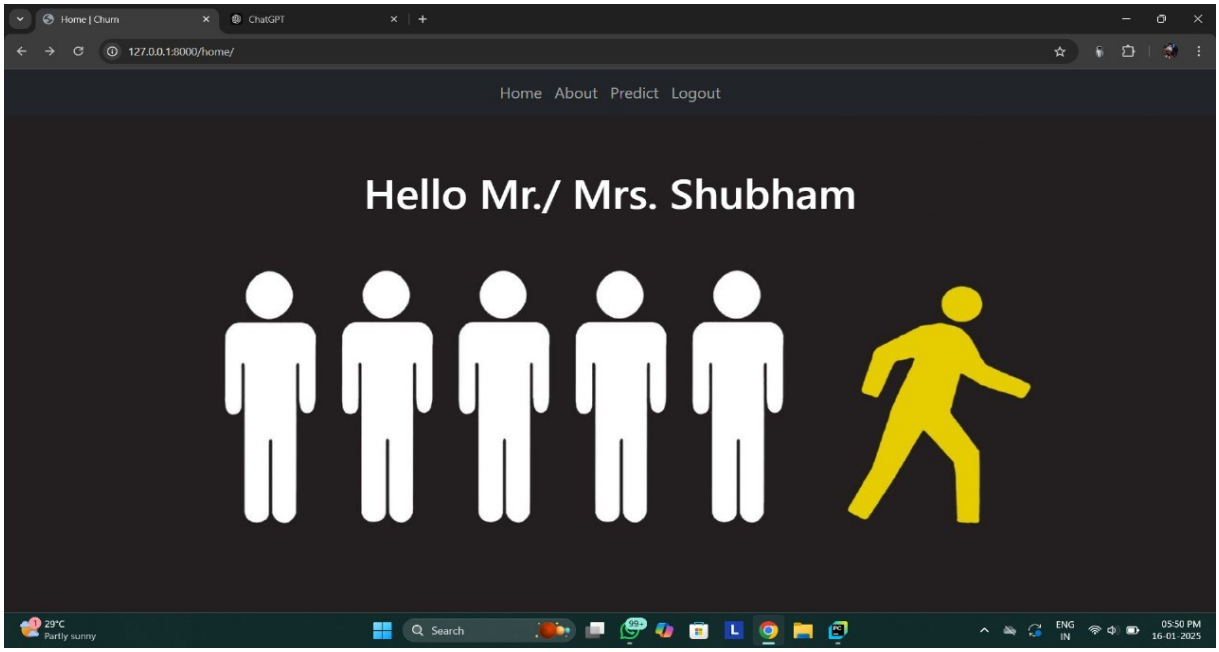

Figure 5: Login Page

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double,Blind Peer,Reviewed, Open,Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

About Page:

The screenshot displays the about page, it shows the information about Customer Churn Prediction.

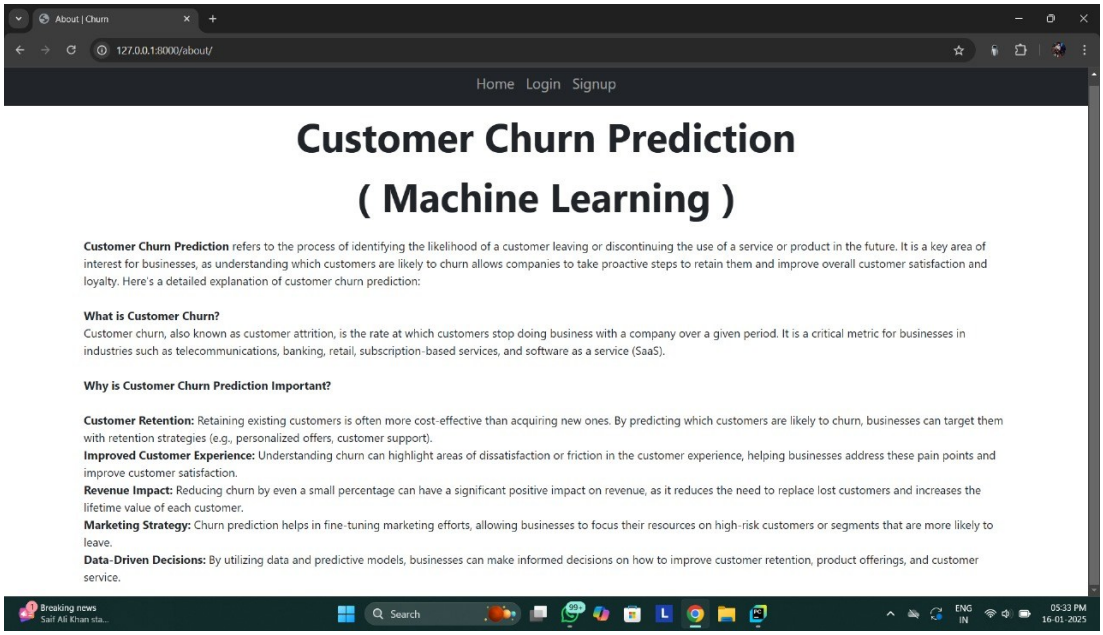

Figure 6: About Page

VI. CONCLUSION

As we continue to develop our Customer Churn Prediction model, we are focused on refining our approach to ensure it meets the needs of our intended users.Currently, we are analysing various algorithms required for predicting valuable insights into customer behaviors, which can be crucial for retention strategies.Once completed, we anticipate that our model will be a powerful tool for organizations to proactively manage customer relationships and reduce churnoutcomes.

VII. FUTURE WORK

1. **Incorporating Real-time Data:**
   Future iterations could leverage real-time customer data streams to make dynamic, immediate predictions. This could allow businesses to react faster to signs of churn before it happens, rather than relying on historical data alone.
2. **Customer Lifetime Value (CLV) Prediction:**
   Enhancing the project to predict customer lifetime value alongside churn can help prioritize high-value customers. Businesses can then focus retention efforts on customers who bring more long-term value.
3. **Multi-source Data Integration:**
   Integrating external data sources, like social media activity, customer sentiment, or market trends, can enhance the model's understanding of customer behavior beyond just transactional data. This can lead to more accurate churn predictions.

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double,Blind Peer,Reviewed, Open,Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

REFERENCES

[1] Predictive analytics, Retrieved from http://en.wikipedia.org/wiki/Predictive_analytics

[2] Customer attrition, Retrieved from http://en.wikipedia.org/wiki/Customer_attrition

[3] Customer Churn Prediction based on Interpretable ML algorithms. Liwen Ou | Year: 2023 | Conference Paper | Publisher: IEEE https://ieeexplore.ieee.org/document/10549047/

[4] T. Chih-Fong and Y. H. Lu, "Customer churn prediction by hybrid neural networks", Expert Systems with Applications, vol. 36, no. 10, pp. 12547-12553, 2009.

[5] Y. Qiu et al., "Clustering Analysis for Silent Telecom Customers Based on K-means++", 2020 IEEE 4th Information Technology Networking Electronic and Automation Control Conference (ITNEC), vol. 1, 2020.

[6] W. Verbeke et al., "Building comprehensible customer churn prediction models with advanced rule induction techniques", Expert systems with applications, vol. 38, no. 3, pp. 2354-2364, 2011.

[7] V. Umayaparvathi and K. Iyakutti, "A survey on customer churn prediction in telecom industry: Datasets methods and metrics", International Research Journal of Engineering and Technology (IRJET) 3.04, 2016.

[8] H. Bingquan, M. T. Kechadi and B. Buckley, "Customer churn prediction in telecommunications", Expert Systems with Applications, vol. 39, no. 1, pp. 1414-1425, 2012.

[9] Omar Adwan, Hossam Faris, Khalid Jaradat, Osama Harfoushi, Nazeeh Ghatasheh, Predicting customer churn in telecom industry using multilayer preceptron neural networks: modeling and analysis, Life Sci. J. 11 (3) (2014) 75–81.

[10] K. B. Oseman, S.B.M. Shukor, N. A. Haris, F. Bakar, “Data Mining in Churn Analysis Model for Telecommunication Industry”, Journal of Statistical Modeling and Analytics, Vol. 1 No.

[11] T. Mutanen, Customer churn analysis – a case study, Technical Report, Retrieved from, http://www.vtt.fi/ inf/julkaisut/muut/2006/customer_churn_ case_study.pdf, April 12, 2014.

[12] S. V. Nath, Customer Churn Analysis in the Wireless Industry: A Data Mining Approach, Technical Report, retrieved from http://download.oracle.com/owsf_2003/40332.pdf, April 14, 2014.

[13] M. Richeldi and A. Perrucci, “Churn Analysis Case Study”, Technical Report, Telecom Italia Lab, Italy, retrieved from http://sfb876.tu dortmund.de/PublicPublicationFiles/richeldi_perrucci_200 2b.pdf, April 12, 2014.

[14] De, P. P and J. Paulose, "Effective ML Techniques to Predict Customer Churn," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), 2021, pp. 895-902.

[15] Srinivasan, R., Subalalitha, C.N. Sentimental analysis from imbalanced code-mixed data using machine learning approaches. Distrib Parallel Databases https://doi.org/10.1007/s10619-021-07331-4.

[16] K. Goyal, K. Kanishka, K. Vasisth, S. Kansal and R. Srivastava, "Telecom Customer Churn Prediction: A Survey," 2021 3rd International Conference on Advances in Computing,

[17] L. Ning, L. Hua, L. Jie, Z. Guangquan, “A customer churn prediction model in telecom industry using boosting”, IEEE Trans. Ind. Inform. 10 (2014) 1659– 1665.

[18] Sangamnerkar, S., Srinivasan, R., Christhuraj, M.R., Sukumaran, R., “ An ensemble technique to detect fabricated news article using machine learning and natural language processing techniques”, 2020 International Conference for Emerging Technology, INCET 2020, 2020, 9154053

[19] M. D. S. Rahman, M. D. S. Alam and M. D. I. Hosen, "To Predict Customer Churn By Using Different Algorithms," 2022 International Conference on Decision Aid Sciences and Applications (DASA), 2022, pp. 601-604, doi: 10.1109/DASA54658.2022.9765155.

[20] M. D. S. Rahman, M. D. S. Alam and M. D. I. Hosen, "To Predict Customer Churn By Using Different Algorithms," 2022 International Conference on Decision Aid Sciences and Applications (DASA), 2022, pp. 601-604, doi: 10.1109/DASA54658.2022.9765155.

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double,Blind Peer,Reviewed, Open,Access, Multidisciplinary Online Journal

Volume 4 | Issue 2 | April 2025

www.ijirid.in

[21] D. Sikka, Shivansh, R. D and P. M, “Prediction of Delamination Size in Composite Material Using Machine Learning,” 2022 International Conference on Electronics and Renewable Systems (ICEARS), 2022, pp. 1228-1232, doi: 10.1109/ICEARS53579.2022.9752123.

[22] Farquad, H. &Vadlamani, Ravi &Surampudi, Bapi. (2014). Churn Prediction using Comprehensible Support Vector Machine: an Analytical CRM Application. Applied Soft Computing. 19. 10.1016/j.asoc.2014.01.031.

[23] Chih Fong Tsai, “Customer churn prediction through the hybrid neural networks”, Expert Systems with Applications 12764 12534.

[24] R. V. S. Rohit, D. Chandrawat and D. Rajeswari, "Smart Farming Techniques for New Farmers Using Machine Learning", Proceedings of 6th International Conference on Recent Trends in Computing, vol. 177, 2021.

[25] V. Geetha, A. Punitha, A. Nandhini, T. Nandhini, S. Shakila and R. Sushmitha, "Customer Churn Prediction In Telecommunication Industry Using Random Forest Classifier," 2020 International.

