IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 2, March 2024

Enhancing Customer Analytics: A Comprehensive Framework for Effective Churn Prediction

[1]Ms. Anitha R, [2]Mr. Adithyan P S, [3]Ms. Akshaya V A, [4]Mr. Bhuvanesh A

[1]Assistant Professor, Department of Computer Science and Engineering

[2 3 4]Students, Department of Computer Science and Engineering

SRM Valliammai Engineering College, Chennai, Tamil Nadu, India

Abstract: *This paper presents a comprehensive framework designed to forecast churn rates within organizations and provide visual representations for proactive and reactive analyses. Churning is the process of customer discontinuation from a product or service due to dissatisfaction or shifting needs of customers. Since keeping current clients comes with a considerably cheaper cost, acquiring new ones is no longer a smart business plan. By leveraging historical customer data encompassing demographic, behavioral, and transactional attributes, predictive models are developed and evaluated. Given that we already have all of the current customer's data, retention is far more affordable and successful. This paper aims at improving the accuracy and gives a churn prediction model which is suitable for all organizations. There are Five main components in this framework: Data preprocessing, pattern recognition, exploratory data analysis(EDA), Gradient Boost algorithm (ML algorithm to identify whether the customer is satisfied with the service or not) and Churn prediction. The churn prediction will give the results as reactive and proactive analysis. We have used the Gradient Boost algorithm (ensemble learning technique) along with SMOTE-ENN in this paper which gives us an accuracy of 95%.This gradient boost algorithm gives the best accuracy when compared with other machine learning algorithms like, Logistic Regression, Random Forest Classifier and Decision Tree Classifier. Visualization of churned data and risk analysis is facilitated through Power BI.*

Keywords: Gradient Boost, Exploratory Data Analysis, churn prediction, SMOTE-ENN

I. INTRODUCTION

The primary objective of this project is to develop an efficient algorithm capable of effectively classifying churn and no churn customers, aiming to have a better customer retention and to analyze the churn rate to take proactive and reactive measures. People are likely to churn with shifting customer preference and evolution in technological advancements. The cost of retention of existing customers is lower when compared to the acquisition cost. As we have the data of existing customers, we can take proactive steps if they are likely to churn. This churn prediction works both on reactive and proactive measures; reactive analysis comes into play once the churn has happened. Based on the available data, we can come to the conclusion that a particular number of people might churn; this is proactive analysis (precautionary analysis). The entire churn prediction process is divided into five main parts and they are data preprocessing, pattern recognition, machine learning algorithms (Gradient Boost in our case), exploratory data analysis, and churn prediction. Cleaning of data takes place in data preprocessing, it will eliminate all the incomplete, irrelevant, and temporary data from the dataset. Next, it performs pattern recognition and separates the data into numerical and categorical data to perform exploratory data analysis (EDA). Proceeding further for machine learning algorithms, the given data is classified as churn when the customer is likely to unsubscribe from the product or service, and it is considered no churn when they are continuing the subscription. We have made use of SMOTE-ENN (Synthetic minority oversampling technique-edited nearest neighbor), which is a combination of both oversampling and undersampling techniques. This technique plays a crucial role in increasing efficiency. ENN is an extended version of K nearest neighbor. The system predicts the churn rate for input data by analyzing the churn history, and the output is displayed in the frontend with the help of PowerBI (a tool specifically used for data visualization).

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 2, March 2024

II. METHODOLOGY

The methodology for the proposed Customer churn prediction and analytics application would involve the following steps:

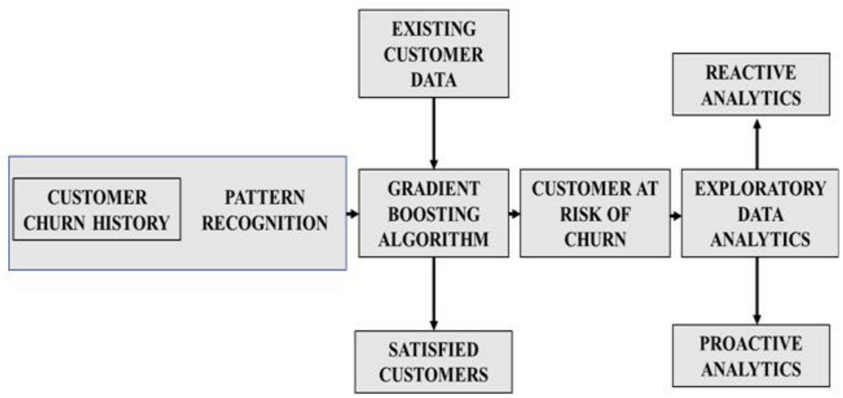

Fig. 1. Architecture diagram of churn prediction framework

1. Data Collection: For this application, data collection from customers involves gathering demographic, behavioral, and transactional information to build comprehensive profiles. This includes capturing customer interactions, preferences, and feedback to feed into the churn prediction model.

2. Data Preprocessing: Data preprocessing involves cleaning, transforming, and preparing raw data for analysis, ensuring its quality and consistency. This stage includes tasks such as handling missing values, removing duplicates, and standardizing data formats to facilitate accurate modeling. Additionally, data preprocessing may involve feature scaling or normalization to ensure all variables contribute equally to the predictive model.

3. Feature Selection: Feature selection is the process of identifying the most relevant variables that contribute to the predictive power of the model. It involves evaluating the importance of each feature and selecting the subset that maximizes predictive accuracy while minimizing complexity.

3. Exploratory Data Analysis: Exploratory data analysis involves visualizing and analyzing data to uncover patterns, trends, and relationships. It helps in understanding the characteristics of the dataset and identifying potential insights that can inform subsequent modeling and decision-making processes.

4. Model training: Model training involves feeding the preprocessed data into various machine learning algorithms to teach it to recognize patterns and relationships within the data. This process iteratively adjusts the model's parameters to minimize prediction errors and optimize its performance for accurately classifying churned and non-churned customers. In this paper Logistic Regression, Decision Tree Classifier, Random Forest Classifier and Gradient Boost Classifier are used.

5. Model Evaluation: Assess the performance of the developed models using appropriate metrics such as Mean Squared Error (MSE) or Root Mean Squared Error (RMSE). Validate the models on a separate test dataset to ensure generalizability.

6. Accuracy Assessment: Quantify the accuracy of the predictive models, considering both training and test datasets. Gradient boost algorithm provides the best accuracy of 95% when compared with other models.

7. SMOTE-ENN: SMOTE-ENN, or Synthetic Minority Over-sampling Technique-Edited Nearest Neighbors, combines oversampling of minority class instances and undersampling of majority class instances to address class imbalance. This technique enhances model performance by generating synthetic data points for the minority class and removing redundant instances from the majority class, improving the classifier's ability to accurately predict rare events like churn.

8. Graphical Representation: Power BI facilitates the creation of visually appealing and interactive dashboards and reports, enabling businesses to effectively visualize churn data and analyze trends, patterns, and insights for informed decision-making.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, March 2024

This methodology enables the creation of an intelligent customer churn prediction and analytics system that utilizes Gradient boost classifier, SMOTE-ENN, exploratory data analysis and data visualization to identify customers who are likely to churn, and empower users with useful insights for informed decision-making.

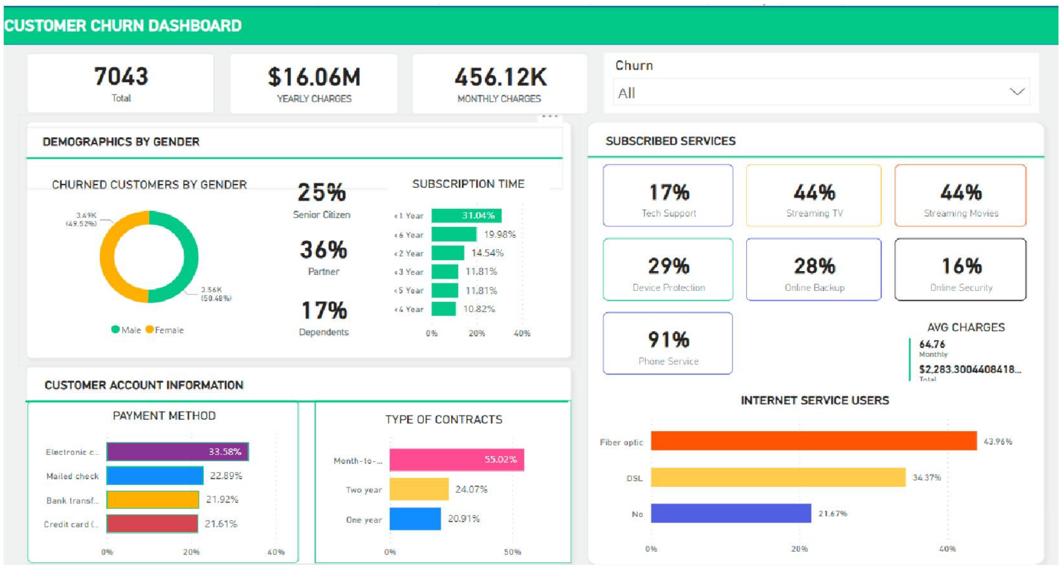

Fig. 2. Graphical representation of churn prediction results

III. RESULTS

The implementation of the churn prediction framework has yielded compelling results, showcasing the efficacy of advanced machine learning techniques in accurately identifying churn patterns within the dataset. Leveraging the Gradient Boost Classifier, the model achieved an impressive accuracy of 95%, signifying its robust performance in distinguishing between churned and non-churned customers. The utilization of SMOTE-ENN played a pivotal role in addressing class imbalance, ensuring that the model effectively learns from both minority and majority class instances, thus enhancing its predictive capabilities.

S.NO	METHODS	ACCURACY(%)	RECALL(%)	F1 SCORE(%)
1.	DECISION TREE	90.80	89	90
2.	RANDOM FOREST	93.26	92	93
3.	GRADIENT BOOST	95.29	96	96
4.	LOGISTIC REGRESSION	90.48	89	90

Fig. 3. Graphical representation of churn prediction results

Furthermore, exploratory data analytics provided invaluable insights into the underlying characteristics of the dataset, revealing key trends and patterns that significantly influence churn behavior. By delving into demographic, behavioral, and transactional attributes, the framework uncovered nuanced relationships and interactions, enabling a more comprehensive understanding of customer churn dynamics.

Moreover, the framework's integration of exploratory data analytics facilitated the discovery of actionable insights that further inform decision-making processes. Through visualizations and data-driven analyses, stakeholders gained a deeper understanding of the factors contributing to churn, enabling them to tailor retention strategies more effectively.

Overall, the results demonstrate the effectiveness of the churn prediction framework in identifying and addressing customer churn, ultimately contributing to enhanced customer retention and sustainable business growth.

IV. CONCLUSION

In conclusion, the development and implementation of a customer churn prediction and analytics framework represent a significant advancement in understanding and mitigating customer attrition. Through the utilization of sophisticated machine learning algorithms, such as Gradient Boosting, coupled with thorough exploratory data analysis, this framework offers invaluable insights into the factors driving churn within a business. The incorporation of clear graphical representations enhances the interpretability and engagement of the findings, providing stakeholders with actionable intelligence to devise targeted retention strategies. By leveraging the power of data-driven decision-making, businesses can proactively identify at-risk customers and deploy personalized interventions to mitigate churn, ultimately fostering long-term customer satisfaction and loyalty. This framework stands as a testament to the transformative potential of analytics in addressing critical challenges within the intricate landscape of customer relationship management, empowering businesses to thrive in today's competitive market environment.

REFERENCES

[1] S. Wu, W. -C. Yau, T. -S. Ong and S. -C. Chong, "Integrated Churn Prediction and Customer Segmentation Framework for Telco Business," in IEEE Access, vol. 9, pp. 62118-62136, 2021, doi: 10.1109/ACCESS.2021.3073776.

[2] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam and S. W. Kim, "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector," in IEEE Access, vol. 7, pp. 60134-60149, 2019.

[3] S. De and P. Prabu, "A Sampling-Based Stack Framework for Imbalanced Learning in Churn Prediction," in IEEE Access, vol. 10, pp. 68017-68028, 2022, doi: 10.1109/ACCESS.2022.3185227.

[4] J. Ahn, J. Hwang, D. Kim, H. Choi and S. Kang, "A Survey on Churn Analysis in Various Business Domains," in IEEE Access, vol. 8, pp. 220816-220839, 2020, doi: 10.1109/ACCESS.2020.3042657.

[5] G. Esteves and J. Mendes-Moreira, "Churn perdiction in the telecom business," 2016 Eleventh International Conference on Digital Information Management (ICDIM), Porto, Portugal, 2016, pp. 254-259, doi: 10.1109/ICDIM.2016.7829775.

[6] N. Jajam, N. P. Challa, K. S. L. Prasanna and C. H. V. S. Deepthi, "Arithmetic Optimization With Ensemble Deep Learning SBLSTM-RNN-IGSA Model for Customer Churn Prediction," in IEEE Access, vol. 11, pp. 93111-93128, 2023, doi: 10.1109/ACCESS.2023.3304669.

[7] A. Amin et al., "Comparing Oversampling Techniques to Handle the Class Imbalance Problem: A Customer Churn Prediction Case Study," in IEEE Access, vol. 4, pp. 7940-7957, 2016, doi: 10.1109/ACCESS.2016.2619719.

[8] E. Lee, B. Kim, S. Kang, B. Kang, Y. Jang and H. K. Kim, "Profit Optimizing Churn Prediction for Long-Term Loyal Customers in Online Games," in IEEE Transactions on Games, vol. 12, no. 1, pp. 41-53, March 2020, doi: 10.1109/TG.2018.2871215.

[9] N. -T. Lee, H. -C. Lee, J. Hsin and S. -H. Fang, "Prediction of Customer Behavior Changing via a Hybrid Approach," in IEEE Open Journal of the Computer Society, vol. 5, pp. 27-38, 2024, doi: 10.1109/OJCS.2023.3336904.

[10] N. Ahmad, M. J. Awan, H. Nobanee, A. M. Zain, A. Naseem and A. Mahmoud, "Customer Personality Analysis for Churn Prediction Using Hybrid Ensemble Models and Class Balancing Techniques," in IEEE Access, vol. 12, pp. 1865-1879, 2024, doi: 10.1109/ACCESS.2023.3334641.

[11]J. Nagaraju, A. S. Sathwik, B. Saiteja, N. P. Challa and B. Naseeba, "Predicting Customer Churn in Insurance Industry Using Big Data and Machine Learning," 2023 First International Conference on Advances in Electrical, Electronics and Computational Intelligence (ICAEECI), Tiruchengode, India, 2023, pp. 1-8, doi: 10.1109/ICAEECI58247.2023.10370876.

[12] S. Menaka, B. R. Ram, N. V. S. Gowtham and G. V. M. Reddy, "Data Transformation, Modelling and Prediction of Customer Churn using Deep Learning," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-6, doi: 10.1109/ICCCNT56998.2023.10306384.

[13] Ş. Dönmez, "Machine Learning-Based Merchant Churn Prediction in Banking," 2023 8th International Conference on Computer Science and Engineering (UBMK), Burdur, Turkiye, 2023, pp. 503-508, doi: 10.1109/UBMK59864.2023.10286753.

[14] J. Vijaya and E. Sivasankar, ‘‘An efficient system for customer churnprediction through particleswarm optimization based feature selectionmodel with simulated annealing,’’ Cluster Comput.,vol. 22, no. S5,pp. 10757–10768, Sep. 2019.

[15] https://www.kaggle.com/blastchar/telco-customer-churn

[16] https://towardsdatascience.com/predict-customer-churn-in-python-e8cd6d3aaa7