JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

ARTIFICIAL INTELLIGENT TOOL USED CHURN PREDICTION

[1]D.Ramesh, [2]M. Sreenivasu, [3]S.Neelima, [4]K.CH.V.Hari, [5]N.Priyanka, [6]SK.Noor Ahmed

[1]Assistant Professor, [2]Associate Professor, [3]UG Student,[4]UG Student,[5]UG Student,[6]UG Student,

Department of Information Technology,
GIET Engineering College, Rajamahendravaram, Andhra pradesh, India

Abstract : **The rise of service providers in various industries has led to increased competition and customer acquisition costs. To sustain their growth, service providers need to focus on keeping their existing clients satisfied and reducing customer churn. Customer churn refers to the phenomenon when customers stop buying from or engaging with a business, and it is a critical issue in industries that rely heavily on customer engagement. The banking and telecommunications sectors, for instance, have recognized the importance of retaining their customer base, and they are investing heavily in predictive analytics to identify customers who are most likely to churn. Machine learning techniques have shown great promise in predicting customer churn.so we used "XGBOOST" algorithm and "LOGISTIC REGRESSION", to identify the most effective model for predicting churn. The aim is to identify the attributes or features that are most strongly associated with churn and develop a predictive model that can identify customers who are at high risk of churning. The application of machine learning techniques for churn prediction is not limited to telecommunications. Other industries such as e-commerce, healthcare, and transportation are also leveraging machine learning to improve customer retention. With the vast amounts of data available, businesses can gain insights into their customers' behavior, preferences, and needs. By leveraging these insights, service providers can personalize their services, improve customer satisfaction, and reduce customer churn.**

Keywords: **Customer Churn, Predictive Analytics, Machine Learning, Xgboost, Logistic Regression, Customer Retention**

1. INTRODUCTION

The telecommunications sector has become one of the main industries in developed countries. The technical progress and the increasing number of operators raised the level of competition. Companies are working hard to survive in this competitive market depending on multiple strategies. Three main strategies have been proposed to generate more revenues: (1) acquire new customers, (2) upsell the existing customers, and (3) increase the retention period of customers. However, comparing these strategies taking the value of return on investment (RoI) of each into account has shown that the third strategy is the most profitable strategy, proving that retaining an existing customer cost much lower than acquiring a new one, in addition to being considered much easier than the upselling strategy. To apply the third strategy, companies have to decrease the potential of customer's churn, known as "the customer movement from one provider to another". Customers' churn is a considerable concern in service sectors with highly competitive services. On the other hand, predicting the customers who are likely to leave the company will represent a potentially large additional revenue source if it is done in the early phase. Many research confirmed that machine learning technology is highly efficient to predict this situation. This technique is applied through learning from previous data

"Churning" is often defined as the percentage of consumers that end their contracts due to competition. Churners are people who have stopped doing business with a firm because they were unhappy with the service they received. An examination of the likelihood of a customer discontinuing use of a service or product is what is meant by a customer churn analysis. Preventive measures are necessary to avoid a customer's items or services being left behind in the case of a circumstance like this.

The marketplace is very dynamic and distinctly aggressive in recent times. It is because of the supply of a big wide variety of service providers. Customers are a company's most valuable asset since they represent its primary source of revenue. Businesses are now cognizant of the fact that they must pay close attention not just to get new clients but also to keep their existing ones satisfied. A churner is a person who moves around a lot and has a variety of reasons for doing so. Customer churn is minimised when the organisation can accurately forecast the customer's mindset and establish linkages between client attrition and things that are under their control. Predicting churn rates is a binary classification job that separates churners from non-churners. For any enterprise, vanquishing commercial enterprise from new customers means going through the sales pipeline, using their sales and advertising belongings in the cycle. Customer retention, then again, is generally extra budget- effective, due to the fact they have already won the confidence and loyalty of current customers. So, predicting customer churn rate at the earlier stages is really important for an organization. In order to avoid the previously described inconvenience, businesses must be able to accurately forecast the purchasing behaviour of their customers. There are two ways to control customer churn: First and foremost, (1) Reactive (2) Proactive. The churners and non-churners are classified in a binary job. In order to deal with this problem, we used the following machine learning techniques: Using Logistic Regression, KNN, Decision Tree classifier, Random Forest Classifier.

2. Literature Survey

Many approaches were applied to predict churn in telecom companies. Most of these approaches have used machine learning and data mining. The majority of related work focused on applying only one method of data mining to extract knowledge, and the others focused on comparing several strategies to predict churn.

[1]. D. Buo and M. Kjellander, 'Predicting Churn Rates at a Swedish CRM-System Company, Linkopings,Universitet,' Linkoping, 2014. In this study, The authors have used logistic regression and decision tree algorithms to predict customer churn rates. They have evaluated the performance of these algorithms using various metrics such as accuracy, precision, recall, and F1 score. The authors have also discussed the various factors that contribute to customer churn, such as customer satisfaction, service quality, and customer loyalty by using machine learning and data mining techniques to predict customer churn. The thesis also covers the data collection and analysis methodology used by the authors to predict customer churn. The authors have collected data from the company's customer database and used statistical analysis techniques to identify the key predictors of churn.[4]. N. Gordini and V. Veglio, 'Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in the B2B e-commerce industry,' Industrial Marketing Management, vol. 62, pp. 100-107, 2017. In this study they developed a churn prediction model tailored for the B2B e-commerce industry by testing the forecasting capability of a new model, the support vector machine (SVM) based on the AUC parameter-selection technique (SVMauc). The predictive performance of SVMauc is benchmarked to logistic regression, neural network and classic support vector machine. Our study shows that the parameter optimization procedure plays an important role in the predictive performance and the SVMauc points out good generalization performance when applied to noisy, imbalance and nonlinear marketing data outperforming the other methods. [5]. 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,' IEEE Access, pp.60134-60149, 2019. In this study, The authors have used the random forest algorithm to predict customer churn rates. They have evaluated the performance of the algorithm using various metrics such as accuracy, precision, recall, and F1 score. The authors have also compared the performance of the random forest algorithm with other machine learning techniques for churn prediction. The article also includes a factor analysis to identify the most significant factors that contribute to customer churn in the telecom sector. The authors have used the random forest algorithm to identify the key factors that affect customer churn, and they have ranked these factors based on their importance. The authors have collected data from a large telecom company's customer database and used statistical analysis techniques to identify the key predictors of churn. [9]. de Lima Lemos, R.A., Silva, T.C. & Tabak, B.M. Propension to customer churn in a financial institution: a machine learning approach. Neural Comput & Applic 34, 11751–11768 (2022). This article presents a study that aimed to evaluate supervised classifiers used in banking to predict customer churn. The study used a unique and representative dataset of a large Brazilian bank at the customer level over time. The authors conducted a horse race of supervised learning classification algorithms under the same validation and evaluation methodology to determine the algorithm that is best suited for the dataset. The study found that the random forests model achieved the best results, even compared to an ensemble model composed of other classifiers. The authors also identified attributes with the highest predictive power of a potential customer churn, which included the frequency with which customers used financial services, the volume of credit extended to them, and their possession of products. The study concludes that machine learning can help banks understand their customers' behavior in an automated way, thereby enabling them to act proactively to reverse potential customer churn and mitigate revenue losses. The random forests model yielded consistent models with superior results, and 80.2% of customers who would churn in the following months were identified. The study highlights the potential for using predictive modeling to add value to banks' customer retention efforts.

3. Overview of the System

3.1 Existing system

The described research aims to predict customer churn using various machine learning algorithms such as Logistic Regression, K Nearest Neighbors (KNN), Decision Tree classifier, Random Forest classifier, Extra Tree classifier, Ridge classifier, Bagging Ridge classifier, and Gradient Boosting algorithms. The study utilizes data from banks or telecom databases and focuses on evaluating the level of customer dissatisfaction primarily through two main indicators: the length of customer association and customer complaints. The choice of algorithms provides a diverse set of techniques for modeling churn prediction, each with its own strengths and weaknesses. Logistic Regression is a simple yet effective method for binary classification tasks, while KNN leverages similarity metrics to identify patterns in the data. Decision trees offer interpretability and can handle non-linear relationships, while ensemble methods like Random Forest and Gradient Boosting enhance predictive accuracy by combining multiple weak learners. The features selected for evaluation, namely service failure rate, length of customer association, and customer complaints, reflect key aspects of customer behavior and experience that are likely to influence churn. By focusing on these variables, the researchers attempt to address limitations in the available data and capture indicators of dissatisfaction that are feasible to measure within the context of the database. Feature extraction and selection play a crucial role in optimizing model performance and interpretability. Through a systematic process, the authors identify relevant features that contribute most significantly to predicting churn. This step helps reduce dimensionality and enhance the efficiency of the modeling process. The research represents a comprehensive effort to leverage machine learning techniques for predicting customer churn in the banking or telecom sector. By considering a range of algorithms and focusing on pertinent features, the study provides insights that can inform strategies for customer retention and satisfaction management within these industries.

3.2 Proposed system

This paper focuses on evaluating the effectiveness of two popular machine learning algorithms, namely Logistic Regression and XGBoost, for predicting customer churn. The study is conducted on a dataset sourced from Kaggle repository and the accuracy, precision, recall, specificity, and False Positive Rate of each algorithm are compared to determine the most efficient one. The experiments revealed that both Logistic Regression and XGBoost are good algorithms for churn prediction, with XGBoost having a slight edge over Logistic Regression due to its speed, stability, and robustness to randomness. XGBoost is a boosting algorithm that uses a group of weaker trees to predict the target variable more accurately. It is well-suited for tabular data and provides efficient inference. XGBoost also has a shorter optimization time for hyper-parameter tuning, making it a practical choice for large datasets. In contrast, Logistic Regression is an extension of linear regression that predicts the probability of a customer belonging to one group or another. It is based on probabilistic calculations and identifies relationships between the target feature and other features. The high accuracy achieved by both algorithms highlights the potential of machine learning in predicting customer churn in a business context. The results of this study suggest that XGBoost can be a preferable choice over Logistic Regression, especially for large and complex datasets. The findings also emphasize the importance of considering different machine learning algorithms and their strengths and limitations before selecting the most appropriate one for a specific application. The logistic regression is essentially an extension of a linear regression, only the predicted outcome value is between [0, 1]. The model will identify relationships between our target feature, Churn, and our remaining features to apply probabilistic calculations for determining which class the customer should belong to.

- Accuracy is high.
- Efficient inference
- Rate of hyper-parameter tuning (shorter the optimization time, the better).

3.3 Software Requirements:

Python
Pandas
NumPy.
Sklearn library
VSCode

3.4 Hardware Requirements:

- Processor: Core i5
- RAM: 4 GB
- OS: Windows 7/8/10 (32 or 64 bit)

GoogleColab

4. Technologies Used

4.1 PYTHON:

Python is a high-level, versatile programming language known for its simplicity, readability, and flexibility. Created by Guido van Rossum and first released in 1991, Python has grown to become one of the most popular languages worldwide, powering a vast array of applications across various domains, from web development and data science to artificial intelligence and automation.

1. Simple and Readable Syntax:

Python's syntax is designed to be intuitive and readable, resembling plain English, which makes it accessible to beginners and experienced programmers alike. Its use of indentation for code blocks instead of curly braces or keywords enhances readability and encourages good coding practices.

2. Versatility:

Python's versatility is one of its defining features. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming, allowing developers to choose the most suitable approach for their projects. This versatility makes Python suitable for a wide range of applications, from small scripts to large-scale software systems.

3. Extensive Standard Library:

Python comes with an extensive standard library that provides modules and packages for performing a wide variety of tasks, such as file I/O, networking, data manipulation, and web development. This rich set of libraries accelerates development by providing pre-built solutions to common programming challenges, reducing the need for developers to write code from scratch.

4. Third-Party Libraries and Ecosystem:

In addition to its standard library, Python boasts a vibrant ecosystem of third-party libraries and frameworks developed by the community. These libraries cover a wide range of domains, including web development (e.g., Django, Flask), data science (e.g., NumPy, pandas, scikit-learn), machine learning (e.g., TensorFlow, PyTorch), and more. The availability of these libraries extends Python's capabilities and accelerates development in specialized areas.

4.2 GoogleColab

Google Colab, short for Google Colaboratory, is a cloud-based platform provided by Google that allows users to write, run, and share Python code directly in their web browsers. It provides an environment for coding, executing code, and collaborating with others in real-time. Google Colab is particularly popular among data scientists, machine learning engineers, educators, and researchers due to its seamless integration with popular machine learning frameworks and libraries like TensorFlow, PyTorch, and scikit-learn. One of the key features of Google Colab is its provision of free access to computing resources, including CPU, GPU, and even TPU (Tensor Processing Unit). Users can leverage these resources without the need for expensive hardware or setup, making it accessible to a wide range of users, including students and hobbyists. This accessibility democratizes computational resources and enables users to tackle computationally intensive tasks such as training deep learning models or analyzing large datasets. Google Colab provides a Jupyter notebook environment, allowing users to create interactive documents that combine code, visualizations, explanatory text, and multimedia elements. This interactive nature facilitates data exploration, experimentation, and collaboration. Users can write Python code in cells and execute them individually or sequentially. The results, including plots, tables, and output messages, are displayed inline within the notebook, making it easy to analyze and interpret the results.

4.3 Libraries

Pandas and NumPy are two powerful Python libraries widely used in data analysis, scientific computing, and machine learning. They offer efficient tools for handling and manipulating large datasets, performing mathematical operations, and conducting data exploration and manipulation tasks. NumPy, short for Numerical Python, is a fundamental package for numerical computing in Python. It provides support for multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays efficiently. NumPy's array objects, known as nd-array, allow for efficient storage and manipulation of large datasets, making it the backbone of many scientific computing tasks.

4.4 System Design:

Designing a churn prediction system involves several components and considerations to ensure its effectiveness and scalability. Below is an outline of the system design for a churn prediction system:

- Data Collection - Identify relevant data sources such as customer transaction history, demographic information, service usage data, customer interactions, and feedback. Establish data pipelines to ingest and process data from various sources, ensuring data quality and consistency. Utilize batch processing or real-time streaming for continuous data ingestion and updates.
- Data Preprocessing and Feature Engineering - Perform data cleaning, handling missing values, and outliers to ensure data quality. Conduct exploratory data analysis (EDA) to understand data distributions, correlations, and patterns. Engineer relevant features such as customer tenure, frequency of interactions, service usage metrics, customer satisfaction scores, and demographic information. Normalize or scale numerical features and encode categorical variables as necessary.
- Model Development - Select appropriate machine learning algorithms such as logistic regression, decision trees, random forests, gradient boosting machines, or neural networks based on the problem requirements and data characteristics. Split the dataset into training, validation, and test sets for model evaluation. Train multiple models using different algorithms and hyperparameter configurations to compare their performance. Optimize hyperparameters using techniques like grid search, random search, or Bayesian optimization to improve model performance.
- Model Evaluation - Evaluate model performance using appropriate metrics such as accuracy, precision, recall, F1-score, ROC-AUC, and lift. Assess model generalization using cross-validation techniques to ensure robustness across different data subsets. Conduct model interpretation and feature importance analysis to understand the factors driving churn predictions.
- Deployment - Deploy the trained model into a production environment using containerization technologies like Docker or serverless platforms like AWS Lambda. Implement model monitoring and logging mechanisms to track model performance, data drift, and concept drift over time. Establish APIs or endpoints for integrating the churn prediction model with existing business systems or applications.
- Feedback Loop and Model Refinement - Monitor the deployed model's performance in production and collect feedback from users or stakeholders.Continuously update and retrain the model using new data to adapt to evolving customer behavior and changing business dynamics. Incorporate user feedback and domain knowledge to refine the model's predictions and improve its accuracy and relevance.
- Documentation and Governance - Document the entire churn prediction system including data sources, preprocessing steps, model architecture, evaluation metrics, and deployment procedures. Establish governance policies and procedures for data privacy, security, and regulatory compliance. Ensure transparency and accountability in model development, deployment, and maintenance processes.

Data Gathering: •

- Customer demographics: This involves collecting information about customers' characteristics such as age, gender, location, occupation, etc. This data can be gathered through customer registration forms, surveys, or from existing databases.
- Usage patterns: Data on how customers use your products or services, including frequency of usage,duration of usage, features utilized, etc., can be collected through analytics tools, user activity logs, telemetry data from products.
- Customer support interactions: Gathering data on customer support interactions involves collecting information about customer inquiries, complaints, feedback, and resolutions. This data can be obtained from customer support tickets, call center logs, chat transcripts, or email communications.
- Historical churn events: Historical churn events refer to instances where customers have discontinued their relationship with your company or stopped using your products/services. This data can be gathered from churn records, cancellation requests, or subscription end dates

5. Architecture

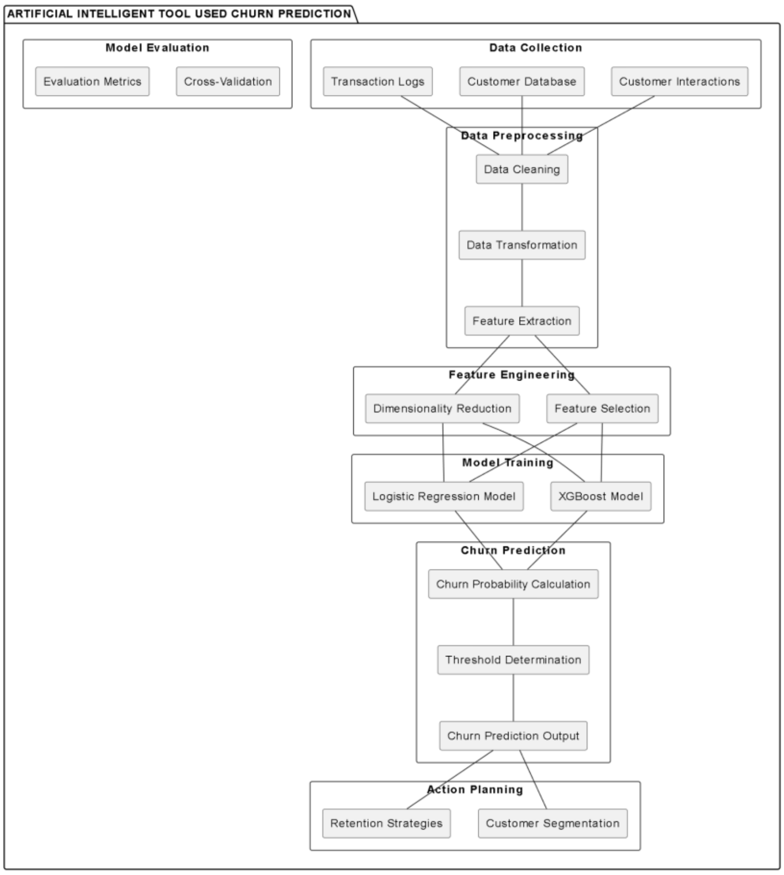

Fig 1: System Architecture

Economical Feasibility:

Cost-effectiveness: Implementing machine learning techniques for churn prediction can be economically feasible due to the potential for significant cost savings associated with retaining existing customers compared to acquiring new ones. By accurately identifying customers at high risk of churning, businesses can allocate resources more efficiently towards targeted retention efforts, thereby reducing overall churn-related costs. Return on Investment (ROI): The investment in implementing machine learning models for churn prediction can yield substantial returns through improved customer retention rates and increased customer lifetime value. Businesses can justify the initial investment by assessing the potential ROI in terms of reduced customer acquisition costs and increased revenue from retained customers.

Technical Feasibility:

Data Availability: The technical feasibility relies on the availability of relevant data sources, such as customer transaction history, demographic information, and engagement metrics. With the rise of big data technologies and advanced analytics platforms, accessing and processing large volumes of data for churn prediction has become more feasible for businesses across various industries. Algorithm Selection: XGBoost and Logistic Regression are widely adopted machine learning algorithms for churn prediction due to their effectiveness and scalability. Their implementation is technically feasible, especially with the availability of libraries and frameworks for machine learning in programming languages like Python and R.

Social Feasibility:

Customer Privacy and Ethics: Businesses must ensure that the implementation of machine learning techniques for churn prediction complies with data privacy regulations and ethical considerations. Respecting customer privacy and maintaining transparency about data usage can enhance social acceptance and trust in the predictive models. Personalization and Customer Satisfaction: Leveraging machine learning insights to personalize services and improve customer satisfaction can enhance social feasibility by fostering positive customer experiences and loyalty. Customers may appreciate personalized recommendations and tailored communications based on their preferences and behavior, leading to stronger social connections with the brand. The proposed system demonstrates economic feasibility by potentially reducing churn-related costs, technical feasibility through available data and algorithmic choices, and social feasibility by addressing privacy concerns and enhancing customer satisfaction. However, businesses need to carefully consider these factors and address potential challenges to ensure successful implementation and adoption of machine learning for churn prediction across various industries.

6. RESULTS SCREENSHOTS

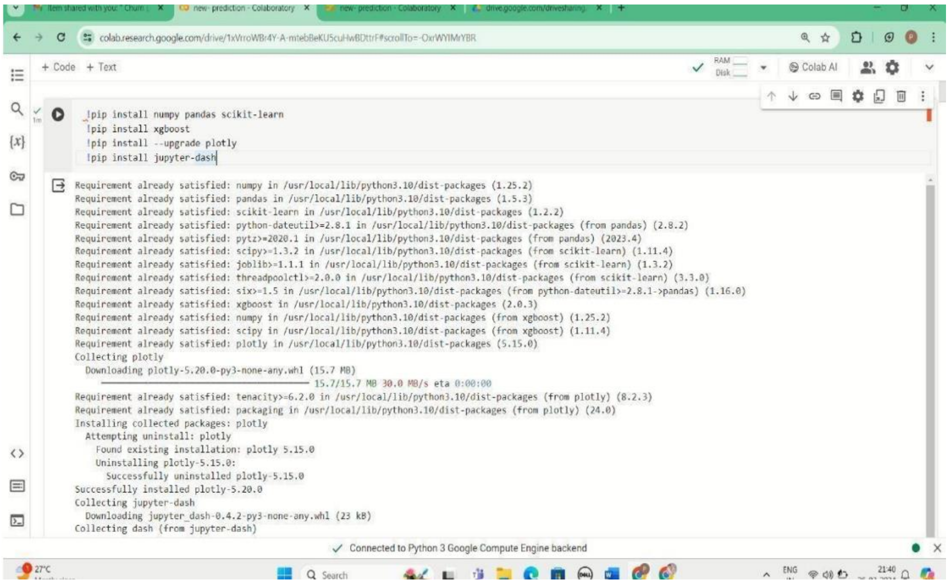

Fig 2: Installation

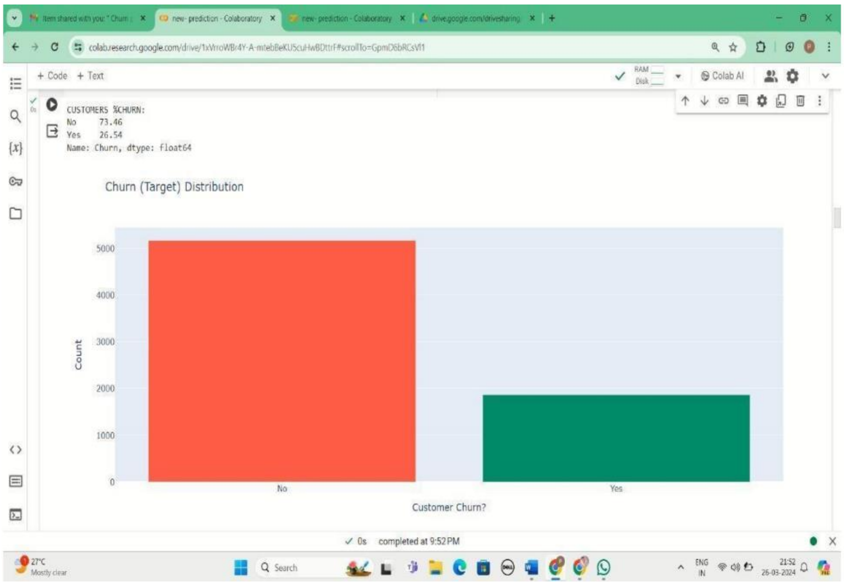

Fig 3: Churn(target)

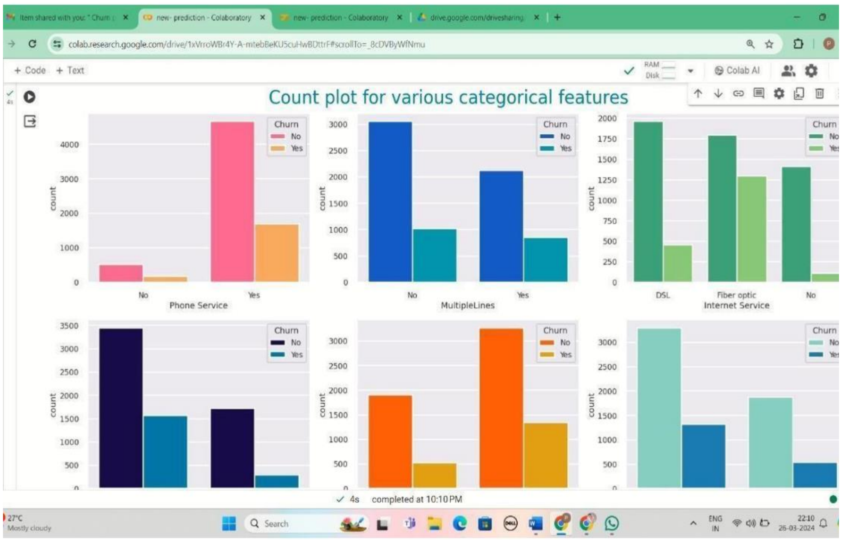

Fig 4:Count Plot

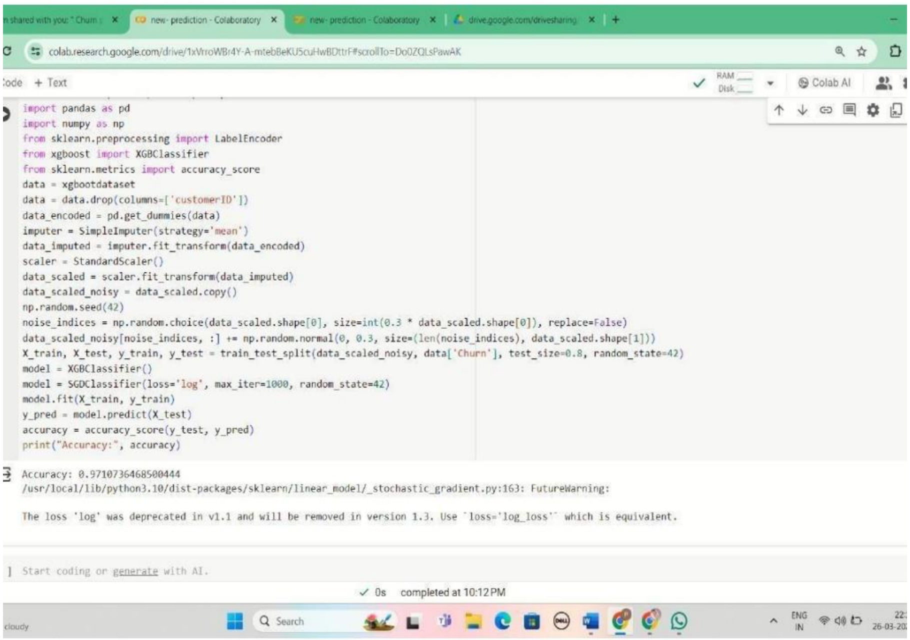

Fig5: Final Output

7. Conclusion

The telecommunications industry grapples with high churn rates, leading to substantial revenue losses. However, effective churn management strategies can mitigate these losses and maintain churn at acceptable levels. This research is critical for telecom companies as predicting churn stands out as a primary driver for profitability. By implementing advanced methods, companies can anticipate customer churn and take proactive measures to retain valuable customers. In this study, the data was divided into training and testing sets with an 80-20 split, ensuring robust model evaluation. Feature engineering techniques were applied to enhance the predictive power of the model, including effective feature transformation and selection methods. Moreover, the issue of imbalanced data, where only 5% of entries represented customer churn, was addressed. Techniques such as under-sampling or utilizing tree-based algorithms, which are resilient to imbalanced data, were employed to overcome this challenge. The results highlighted the effectiveness of logistic regression and XGBoost tree models in predicting churn. Both models exhibited superior performance across various evaluation metrics, with logistic regression achieving an AUC value of 81.30% and XGBoost tree model achieving 81.102%. These high AUC values signify the models' ability to accurately distinguish between churn and non churn instances. The findings underscore the importance of employing sophisticated machine learning techniques, such as logistic regression and XGBoost, for churn prediction in the telecommunications sector. By leveraging these models, telecom companies can identify customers at high risk of churn and implement targeted retention strategies to retain them. Ultimately, this approach can lead to enhanced customer satisfaction, reduced churn rates, and increased profitability for telecom operators. This research contributes valuable insights into churn prediction methodologies and underscores the significance of advanced analytics in driving business growth and sustainability in the telecommunications industry. By leveraging predictive analytics effectively, telecom companies can proactively address churn challenges and position themselves for long-term success in a competitive market landscape.

8. Future Enhancement

The future scope of this paper encompasses the integration of hybrid classification techniques to explore the correlation between churn prediction and customer lifetime value (CLV) in the telecommunications industry. By leveraging hybrid classification models, which combine multiple algorithms or methodologies, researchers can uncover intricate patterns and relationships that may not be discernible with individual models alone. This approach will enable a deeper understanding of the factors influencing both churn prediction and CLV, ultimately leading to more effective customer retention strategies. Furthermore, the future research direction involves incorporating retention policies into the churn prediction framework. By identifying and selecting appropriate variables from the dataset that are indicative of customer retention strategies, researchers can enhance the accuracy and relevance of churn prediction models. This integration will enable telecom companies to proactively address churn risk factors and implement targeted retention initiatives tailored to specific customer segments. The passive and dynamic nature of the telecommunications industry underscores the significance of data mining and predictive analytics in shaping its future prospects. As the industry continues to evolve, the volume and complexity of data generated from various sources such as customer interactions, service usage patterns, and market trends will increase exponentially. Consequently, data mining techniques will play an increasingly vital role in extracting actionable insights from these vast datasets, driving informed decision-making and strategic planning in telecom companies.

9. References

Paper References:

[1] A. K. Ahmad, A. Jafar, and K. Aljoumaa, "Customer churnprediction in telecom using machine learning in big data platform," Journal of Big Data, vol. 6, no. 1, pp. 28–24, 2019.

[2] D. Paulraj, "An automated exploring and learning model fordata prediction using balanced CASVM," Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 5, 2020.

[3] C. Saravanakumar, R. Priscilla, B. Prabha, A. Kavitha,M. Prakash, and C. Arun, "An efficient on-demand virtual machine migration in cloud using common deployment model," Computer Systems Science and Engineering, vol. 42, no. 1, pp. 245–256, 2022.

[4] H. Jain, A. Khunteta, and S. Srivastava, "Churn prediction in telecommunication using logistic regression and logit boost,"Procedia Computer Science, vol. 167, pp. 101–112, 2020.

[5] T. Xu, Y. Ma, and K. Kim, "Telecom churn prediction system based on ensemble learning using feature grouping," Applied Sciences, vol. 11, no. 11, p. 4742, 2021.

[6] J. Vijaya and E. Sivasankar, "Computing efficient features using rough set theory combined with ensemble classification techniques to improve the customer churprediction in telecommunication sector," Computing, vol. 100, no. 8,pp. 839–860, 2018.

[7] D. Rb, Customer Churn prediction in telecommunication industry through machine learning based fine-tuned XGBoostalgorithm, in Proceedings of the International Conference on Innovative Computing & Communication (ICICC), Delhi, India, April 2021.

Text Book References:

[1]. D. Buo and M. Kjellander, 'Predicting Churn Rates at a Swedish CRM-System Company, Linkopings,Universitet,' Linkoping, 2014.

[2]. H. Gebert, M. Geib, L. Kolbe and W. Brenner,'Knowledge-enabled customer relationship management: integrating customer relationship management and knowledge management concepts,' Journal of Knowledge Management, vol.7, no. 5, pp. 107-123, 2003.

[3]. F. Khodakarami and Y. Chan, 'Exploring the role of customer relationship management (CRM) systems in customer knowledge creation,'Information & Management, vol. 51, pp. 27-42,2014.

[4]. N. Gordini and V. Veglio, 'Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B Artificial Intelligent Tool Used churn prediction GIET ENGINEERING COLLEGE PAGE 67 ecommerce industry,' Industrial Marketing Management, vol. 62, pp. 100-107, 2017.

[5]. 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,' IEEE Access, pp.60134-60149, 2019.

[6]. A. T. Jahromi, S. Stakhovych and M. Ewing, 'Managing B2B customer churn, retention and profitability,' Industrial Marketing Management, vol. 43, no. 7, pp. 1258-1268, 2014.

[7]. 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,' IEEE Access, pp.60134-60149, 2019.

[8]. K. Coussement and D. Van den Poel, 'Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques,' Expert Systems with Applications, vol. 34, no. 1, pp. 313-327,2008.

[9]. de Lima Lemos, R.A., Silva, T.C. & Tabak, B.M. Propension to customer churn in a financial institution: a machine learning approach. Neural Comput & Applic 34, 11751–11768 (2022)