CUSTOMER CHURN PREDICTION

Mr.R.Phadtare[1], Harsh Shilimkar[2], Sumeet Survase[3], Harshal Patil[4], Gaurav Khedkar[5]

[1]Professor, [2,3,4,5] BE Students, Department of Computer Engineering, SKNSITS, Lonavala, Maharshtra, India. [1]rsphadtare.sknsits@sinhgad.edu, [2]harshs.sknsits.comp@gmail.com, [3]harshalpatil.sknsits.comp@gmail.com, [4]sumitsurvase21@gmail.com, [5]gauravkhedkar1505@gmail.com

ABSTRACT - Our work focuses on applying supervised machine learning techniques to anticipate bank client attrition. We created a straightforward and effective strategy that works with a variety of machine learning techniques. We tested many methods through tests and discovered that the Decision Tree strategy was the most effective in forecasting client attrition. By demonstrating how well supervised machine learning works to handle churn prediction issues in the banking industry, this study advances the discipline.

Keywords: Bank customer churn prediction, Supervised machine learning, Decision Tree approach, Model evaluation, Predictive analytics, Customer retention, Banking sector, Classification algorithms, Experimentation, Model performance.

I. INTRODUCTION

In the banking sector, predicting bank customer churn is a crucial activity that identifies clients who are most likely to quit the bank in the future. Banks now have sophisticated tools to accurately forecast and avoid client attrition thanks to the development of machine learning (ML) techniques, notably those driven by TensorFlow and Kera's. Neural network implementation in data modelling is made possible by the widely used deep learning platforms TensorFlow and Kera's. These frameworks facilitate the development of sophisticated neural network designs that are capable of accurately predicting customer attrition by capturing complicated patterns in consumer data. Several additional Python libraries are essential to the creation and implementation of churn prediction models, in addition to TensorFlow and Kera's. Pandas is used for preprocessing and data manipulation, making sure the data is organized and clean enough to train machine learning models. For data visualization, Matplotlib is used, which makes it easier to grasp model outcomes and insights. By offering tools for a range of machine learning activities, including data preparation, model assessment, and hyperparameter tweaking, Scikit-Learn is a useful addition to TensorFlow and Kera's. In addition to TensorFlow and Kera's, many more Python libraries are necessary for the development and use of churn prediction models. Pandas is used to ensure that data is clean and well-organized enough to train machine learning models through preprocessing and modification. Matplotlib is used for data visualization, which facilitates understanding of model results and insights. Together with TensorFlow and Kera's, Scikit-Learn provides tools for a variety of machine learning tasks, such as data preparation, model evaluation, and hyperparameter tinkering.

II. LITERATURE SURVEY

Due to its vital role in preserving customer loyalty and corporate sustainability, customer churn prediction has attracted a lot of interest in academic research and industrial operations. Numerous research projects have examined various approaches and strategies to accurately forecast and reduce client attrition.

1.Conventional Statistical Methods: Statistical techniques including survival analysis, decision trees, and logistic regression were often used in early customer churn prediction studies. To forecast future customer turnover behaviour, these strategies concentrated on finding patterns and trends in previous customer data.

2. Techniques for Machine Learning: As machine learning gained traction, researchers started looking at increasingly complex algorithms for predicting client attrition. Because supervised learning algorithms can handle complicated data patterns and increase prediction accuracy, they have become extensively used.

3. Feature engineering: By identifying and modifying pertinent characteristics from unprocessed data, feature engineering is essential to the prediction of customer attrition. To enhance the effectiveness of churn prediction models, studies have looked at a variety of feature selection strategies, dimensionality reduction techniques, and domain-specific feature engineering methodologies.

4.Ensemble Learning: To improve the churn prediction models' resilience and capacity for generalization, ensemble

 DOI : 10.35291/2454-9150.2024.0154

learning strategies including bagging, boosting, and stacking have been studied

5. Temporal Analysis: Time-series forecasting, and sequence modelling are two methods of temporal analysis of consumer behaviours that have shown promise in predicting customer attrition.

6.Deep Learning: Because deep learning techniques, especially deep neural networks, can automatically extract hierarchical characteristics from raw data, they have become popular in the prediction of customer turnover. Research has looked at several architectures, such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs), to better forecast outcomes and identify intricate patterns in consumer data.

III. PROBLEM STATEMENT

Our focus is on helping firms anticipate loss of clients and take proactive measures to retain their clientele. Its profits and development can be negatively impacted by customer turnover, so it's critical to identify and keep in danger clients. Our goal is to create techniques that will enable organizations to retain happy and devoted consumers by precisely predicting turnover via the use of analysis and neural networks.

IV. PROBLEM DESCRIPTION

Assume you own a corporation, such as a cell phone provider or video service. a few of your clients sometimes discontinue utilizing your facility. We refer to this as "churn." Drift is a concern as it can result in financial loss and possibly a lag below rivals. Predicting loss of clients can help with that. It's like gazing into a ball of stones, except rather than using magic, we predict whose consumers are most likely to depart using data and intelligent computer algorithms. Businesses may intervene early to prevent clients from departing by doing this. Therefore, the objective is to develop technologies that can precisely forecast churn and assist companies in retaining satisfied and loyal clients.

V. REQUIREMENT SPECIFICATION

1. Building machine learning models to forecast client retention is based on Python:

Python is used exclusively to construct the system's prototype backend, with the Anaconda package being used to facilitate development. A wide range of packages for data processing and prediction tasks are available thanks to Python's vast library ecosystem.

2. Library of Machine Learning:

- TensorFlow and Kera's: –Neural network model construction and training are done with TensorFlow and its high-level API, Kera's.
- Pandas: - Pandas helps with the preparation of raw data and is used for data manipulation and analysis.
- Matplotlib: Matplotlib is used for data visualization, which helps to understand model performance and data trends.
- Scikit-Learn is a machine learning toolkit that may be used for a variety of tasks like feature engineering, data preparation, and model validation.

3.WebsiteConstruction:

Flask: The Python backend will be deployed using the Flask web framework. Python developers can create web applications more easily and with greater flexibility because to its compact and modular design, which also makes maintenance simpler.

4. Implementing the Model:

Pre - trained.h5 Models: The web application uses pre-trained models that have been trained by the system for prediction purposes. Neural networks and Kera's are especially utilized in the Python model construction process, which is carried out locally on computers. These pre-trained models must load and run smoothly in the web context, which requires Kera's and other crucial machine learning libraries.

VI. OBJECTIVES

1. Enhancement of Retention: Increasing client retention rates is the main goal of bank customer churn prediction. Proactive steps to keep clients who are likely to leave the bank can be performed to identify them, which will lower attrition and preserve a steady clientele.

2. Customer Satisfaction: By anticipating customer attrition, banks may act against problems that might be causing customers to become dissatisfied. Through comprehension of the elements that contribute to customer attrition, financial institutions can devise tactics to improve client happiness and experience.

3. Risk Mitigation: By anticipating customer churn, banks may reduce the financial risks of losing important clients. Banks may limit revenue loss by allocating resources properly and prioritizing retention efforts to high-value clients who are at danger of leaving.

4. Marketing Optimization: By directing interventions towards customers who are most likely to churn, banks may optimize their marketing efforts thanks to churn prediction. It is made sure that marketing resources are used effectively, which improves campaign efficacy overall and yields higher returns on investment.

5. Product Development: By comprehending the causes of consumer attrition, product developers and innovators can get important insights. Banks may better fulfil the requirements and preferences of their customers by tailoring

 DOI : 10.35291/2454-9150.2024.0154

International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-10, Issue-01, Apr 2024

their goods and services via the analysis of consumer feedback and behaviour patterns.

6. Competitive Advantage: By helping banks to proactively manage client relationships and remain ahead of competition, predictive churn modelling can provide them an advantage over rival banks. Early identification and management of churn risks helps banks stand out from the competition and better maintain their clientele.

7. Improved Financial Performance: The goal of bank customer churn prediction is to raise the organization's financial performance. Banks can raise revenue, lower client acquisition expenses, and improve overall profitability by lowering attrition rates. Thus, the bank's long-term viability and expansion are facilitated.

VII. Architecture

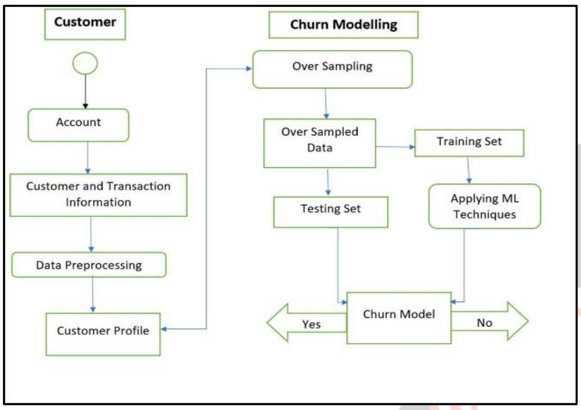

Figure 1: System Architecture

System Architecture Overview:

The architecture of the churn prediction model encompasses several key stages to effectively analyse and predict customer churn. Here's a concise breakdown:

• Data Sources: Customer and Transaction Information: This dataset comprises customer demographics, purchase history, support interactions, and service usage data.

• Account Data: Details about customer accounts, including account type, subscription plan, and payment history, are essential for modelling churn behaviour.

• Data Preprocessing: Data preprocessing involves cleaning and formatting the data, addressing missing values, outliers, and ensuring consistency in data formats. This step is crucial for ensuring the quality of input data for the model.

• Handling Imbalanced Datasets: Imbalanced datasets, where one class (e.g., churning customers) is significantly smaller than the other, are addressed using oversampling techniques. This balances the dataset by replicating data points from the minority class, improving model performance.

• Training and Testing Sets: The pre-processed data is split into training and testing sets. The training set is used to build the churn prediction model, while the testing set evaluates the model's performance on unseen data.

• Machine Learning Model Development: Machine learning techniques are applied to train the model on the training data. The model learns patterns and relationships between customer data and churn behaviour, enabling it to make accurate predictions.

• Customer Profile Generation: Once trained, the model generates a customer profile with a churn probability score for each customer. This score quantifies the likelihood of a customer churning based on their characteristics and behaviour.

• Utilization of Churn Model: Companies leverage the churn prediction model to prioritize outreach efforts and develop targeted retention campaigns for high-risk customers. By focusing resources on customers with a higher likelihood of churning, companies can effectively reduce churn rates and retain valuable customers.

VIII. RESULT

To enhance customer retention rates, the project uses machine learning models to anticipate client churn and provide real-time predictions. Collecting information, preprocessing, analysis of exploratory data, model creation using TensorFlow, Kera's, and Scikit-Learn, and deployment using Flask/Python over the backend & the use of JavaScript, HTML, and CSS for the frontend are just a few of the phases it covers.

• User-Friendly Ui: The front-end development process made sure that an interface that is easy to use was created by utilizing HTML, CSS, and JavaScript. By offering a graphical and easy-to-use interface, this interface improves the user's experience with the program.

• Smooth Data Upload: The capability that enables users to upload Excel files with client data simplifies the churn prediction procedure. By supporting different user types, this feature improves the application's usability and accessibility.

• Prediction Analytics: The online application offers statistical analysis for customer attrition by utilizing machine learning models constructed with Tensor flow, Kera's, Pandas, which is and Scikit-Learn. By identifying intricate patterns and correlations in the data, the integration of various libraries makes precise predictions possible.

• Reason Analysis: The program not only forecasts client attrition but also examines the causes of customer leave. This study provides insightful information on the elements that contribute to employee attrition, enabling companies to create focused retention plans.

• Real-Time Prediction: Real-time customer churn probability prediction is made possible by the web application's pre-loaded.h5 models. This feature enables

 DOI : 10.35291/2454-9150.2024.0154

prompt insights into possible churn scenarios, which supports proactive the decision-making process.

• Scalability and Efficiency: The programming language's backend is deployed using the Flask web framework, which guarantees scalability and efficiency in processing user requests. Because of its flexible and lightweight architecture, Flask is a great choice for Python web application development.

• Actionable information: The program delivers firms actionable information to maximize client retention efforts by offering churn projections as well as reasons for leaving. With the use of these information, firms can effectively customize their tactics to target individual churn reasons.

• Iteration and continual improvement are made possible by the web application's modular architecture. The program may be updated to include the most recent discoveries and improvements when new data becomes accessible and as models are retrained, guaranteeing its relevance and efficacy throughout time.

IX. CONCLUSION

This project aimed to forecast customer churn, also known as attrition, through a systematic process. It began with collecting comprehensive client information, including various metrics, interactions, and historical churn patterns. After gathering the data, rigorous cleaning procedures were applied to handle missing values and ensure data quality. Additionally, feature engineering techniques were employed to enhance the predictive power of the model, focusing on relevant factors such as client tenure and purchase frequency.

Various machine learning techniques, such as logistic regression and decision trees, were evaluated and implemented based on their suitability for the task. These models were trained using a portion of the data to optimize their performance metrics. The performance of the trained models was then evaluated using a range of metrics including reliability, precision, recall, F1-score, and ROC AUC.

The most effective model was subsequently deployed for real-time churn prediction, allowing proactive measures to be taken to retain customers. Continuous monitoring and refinement of the deployed model were carried out to ensure its ongoing accuracy and effectiveness in predicting and mitigating customer churn. Through this comprehensive approach, the project successfully developed and deployed a predictive model for customer churn, enabling proactive customer retention efforts and ultimately contributing to the overall success of the business.

X. REFERENCES

[1] S. 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, doi: 10.1109/ICIRCA51532.2021.9544785.

[2] I.Japparova and R. Rupeika-Apoga, “Banking business models of the digital future: The case of Latvia,” European Research Studies Journal, vol. 20, no. 3, pp. 864–878, 2017, doi: 10.35808/ersj/749.

[3] G. Nie, W. Rowe, L. Zhang, Y. Tian, and Y. Shi, “Credit card churn forecasting by logistic regression and decision tree,” Expert Systems with Applications, vol. 38, no. 12, pp. 15273– 15285, Nov. 2011, doi: 10.1016/j.eswa.2011.06.028.

[4] M.-S. Chen, J. Han, and P. S. Yu, “Data mining: an overview from a database perspective,” IEEE Transactions on Knowledge and data Engineering, vol. 8, no. 6, pp. 866–883, 1996.

[5] Ke, X. and Cheng, J. (2018). Customer attrition prediction in telecommunications utilizing machine learning on a big data platform. Journal of large-scale data, 5(1), 1–17.

[6] Deng S., Han Y., Hu X., & Guo X. (2020). Innovation in large-scale data analytics to forecast customer attrition. IEEE Access

[7] Fumera, G. and Roli, F. (2005). A flexible classifier selection algorithm used in economic time series forecasting. Neurocomputing, 64, 343–362.

[8] Liu F., Wang H., Zhang L., and Wang S. (2019). Temporal convolutional networks can forecast attrition in subscription services. Expert Systems has with Applications, 115, 185–195.

[9] Vellido, A., Lisbon, P. J., and Meehan, K. (1999). An industrial market segmentation comparison using SOM neural networks. Computers and Operations Research, 26(1), 59–77.

 DOI : 10.35291/2454-9150.2024.0154