CUSTOMER CHURN PREDICTION

[1] Gaurav Khedkar, [2] Harshal Patil, [3] Sumeet Survase, [4] Harsh Shilimkar, [5]Prof. Mr.R.Phadtare
SKN Sinhgad Institute of Technology, Lonavala, India

ABSTRACT

The popularity of using Internet contains some risks of network attacks. Intrusion detection is one major research problem in network security, whose aim is to identify unusual access or attacks to secure internal networks. In literature, intrusion detection systems have been approached by various machine learning techniques. In this literature, we propose a real-time intrusion detection approach using a supervised machine learning technique. Our approach is simple and efficient, and can be used with many machine learning techniques. We applied different well-known machine learning techniques to evaluate the performance of our IDS approach. Our experimental results show that the Decision Tree technique can outperform the other techniques. Therefore, we further developed a real-time intrusion detection system (RT-IDS) using the Decision Tree technique to classify on-line network data as normal or attack data.

1.INTRODUCTION

The fintech industry's rapid evolution, coupled with changing regulations and shifting customer preferences, has ushered in a new era of competition, challenging the traditionally dominant position of banks in society (The Economist, 2019). In this era of intensified competition, customer turnover, often referred to as customer churn, poses a significant threat to established financial institutions (De Caigny et al., 2018). Customer churn occurs when customers disengage or sever their ties with a company within a specified timeframe (Colgate et al., 1996). As competition continues to escalate, it becomes imperative for banks to retain existing customers, as doing so proves to be more cost-effective than acquiring new ones, ensuring their continued relevance in society.

Advancements in technology have provided banks with greater access to data, enabling the feasibility of data-driven customer churn analysis. Consequently, there is a growing demand for customer churn analysis, a field that explores a set of characteristics to predict and mitigate customer churn (De Caigny et al., 2018). This heightened demand has spurred the use of predictive modeling, employing statistical learning methods, to anticipate and address customer churn (Ganesh et al., 2000). However, an essential question arises: which statistical learning method is most effective in predicting customer churn?

Verbeke et al. (2012) contend that popular methods for predicting churn probability include logistic regression, Naïve Bayes, and decision trees due to their combination of predictive performance and interpretability. Nevertheless, practical challenges exist with Naïve Bayes and decision trees, necessitating the exploration of alternative methods. This study endeavors to evaluate and analyze the performance of additional methods for predicting customer churn, comparing logistic regression, random forest, and K-nearest neighbour models

The selection of the preferred additional method is contingent on the reliability of its predictions. To 6 assess prediction reliability, we will compare evaluation metrics derived from two cross-validation set approaches. In conclusion, this study seeks to identify the optimal statistical learning method for customer churn prediction and determine which cross-validation set approach yields the most dependable results.

There are many competitors in the banking sector nowadays and those competitors are also ready to provide higher quality and lower prices for the same products and services. So, customers are shifting loyalties from bank to bank. When products or services do occur, customer churn or attrition occurs. There are two kinds of churns like voluntary or involuntary churning. The customer leaves the bank or stops using Involuntary churners which the bank removes users. Voluntary churn arose when customers immediately stopped using products or services. This study is focused on the voluntary churn. To prevent this, the prediction of customer churn is necessary to determine whether customers can discontinue products.

In recent years, data mining has become popular in the research industry and in society as a whole, due to the enormous availability of large amounts of data and the need to turn such data into useful information and knowledge. Using SVM, bank customer churn prediction is developed among data mining techniques, because it is widely used binary classification technique and performs good classification accuracy in small dataset. But it increases time complexity in large data set so that clustering of K-means is used to effectively cluster high volume data before SVM to reduce operating time and improve prediction accuracy.

1.2 Motivation

Customer churn prediction in the banking sector is motivated by several key factors: Customer Retention: Banks aim to retain existing customers as it's more cost-effective than acquiring new ones. Predicting churn allows them to take proactive steps to retain valuable customers. Revenue Protection: Losing customers means losing revenue. Predicting churn helps banks identify at-risk customers and take actions to prevent them from leaving, thereby protecting their revenue streams. Customer Experience: Banks want to provide a positive customer experience. Identifying potential churners allows them to address issues, improve services, and enhance the overall customer experience. Risk Management: Identifying customers likely to churn can help banks assess the potential impact on their portfolio and take measures to mitigate risks associated with customer attrition. Marketing Efficiency: Banks can optimize their marketing efforts by targeting retention campaigns at customers with a higher likelihood of churning, thus improving the cost-effectiveness of their marketing strategies. Competitive Advantage: Predicting churn can give banks a competitive edge by allowing them to respond to customer needs more effectively and keep pace with or surpass competitors. Regulatory Compliance: Some regulations may require banks to monitor and report on customer churn to ensure fair treatment of customers. In summary, customer churn prediction is essential in the banking sector to maintain profitability, enhance customer satisfaction, and stay competitive in a rapidly evolving industry.

1.3 Objectives

- Data Collection and Preprocessing: Gather historical customer data, clean, and preprocess it to create a structured dataset for analysis.
- Define Churn: Clearly define what constitutes "churn" in the context of your project. It could be when a customer closes their account or stops using certain banking services
- Feature Selection: Identify relevant features (e.g., transaction history, account age, customer demographics) that may influence churn and select the most important ones
- Data exploring: Analyze the dataset to understand trends, correlations, and patterns that may indicate reasons for churn
- Model Selection: Choose appropriate machine learning algorithms for prediction. Common choices include logistic regression, decision trees, random forests, or neural networks.
- Data Splitting: Split the dataset into training and testing sets to evaluate model performance.
- Model Training: Train the chosen machine learning model(s) using the training data.
- Hyperparameter Tuning: Optimize the model's hyperparameters to improve its predictive accuracy.
- Interpretability: Ensure the model's results are interpretable, so you can understand why certain customers are predicted to churn.
- Deployment: Deploy the trained model into the bank's systems so it can make real-time predictions on new customer data.
- Monitoring: Continuously monitor the model's performance and retrain it periodically to adapt to changing customer behavior. • Feedback Loop: Establish a feedback loop to collect data on the effectiveness of retention strategies and use this information to improve the model and strategies over time.
- Reporting: Provide regular reports and insights to bank stakeholders on customer churn trends and the impact of retention efforts.

2. LITERATURE REVIEW

Literature review is a text of a scholarly paper, which includes the current knowledge including substantive findings, as well as theoretical and methodological contributions to a particular topic. Literature reviews are secondary sources, and do not report new or original experimental work.

3. PROBLEM STATEMENT

"Develop a predictive model that can effectively identify and anticipate customer churn in the banking industry. Utilize historical customer data, transaction records, and customer interactions to forecast which customers are at risk of leaving the bank. The goal is to proactively address customer retention by identifying potential churners and implementing targeted strategies to reduce attrition, thereby preserving the bank's customer base and revenue." This problem statement sets the context for the task of building a predictive model to address the issue of customer churn in the banking sector, emphasizing the use of data and predictive analytics to mitigate customer attrition.

4. PROJECT REQUIREMENT SPECIFICATION

4.1 Software Requirement Specification

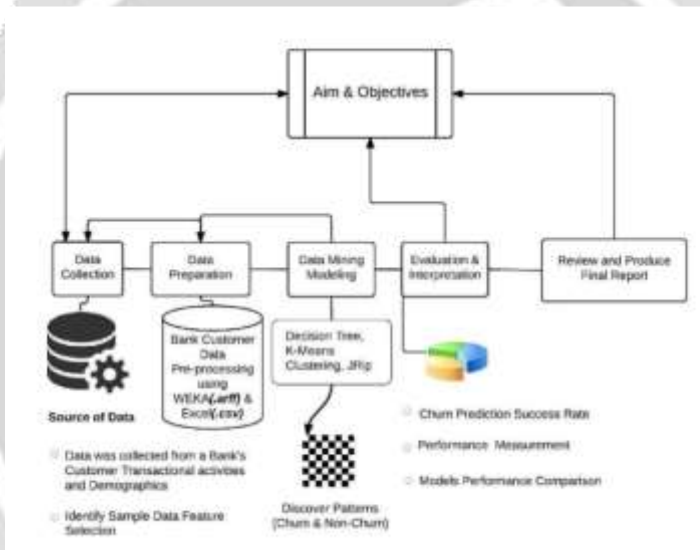
- Operating System: Windows, Mac OS
- Browser: Google Chrome (above 106), Microsoft Edge (107, 108), Safari (above 15.6), Mozilla Firefox (above 106), and Opera (92)

4.2 Hardware Requirements

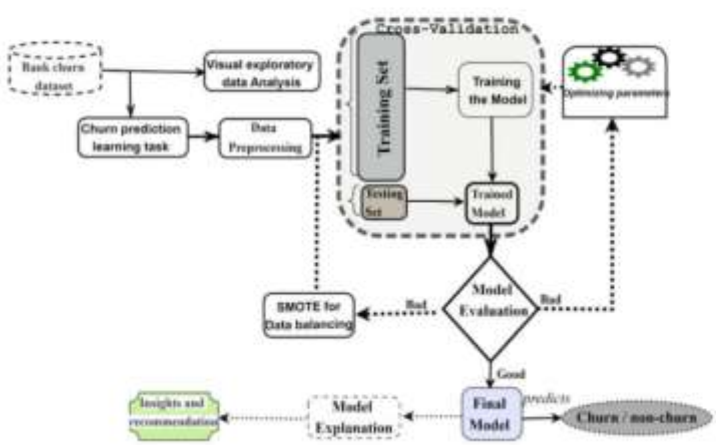
- Processor: i3 or Higher version
- RAM: 2 GB or Higher

5. SYSTEM PROPPOSED ARCHITECTURE

5.1 Architecture Diagram



5.2 Mathematical model



6. Methodologies

- **Data Collection:** Gather historical customer data, including transaction history, demographics, customer service interactions, and any other relevant information.
- **Data Preprocessing:** Clean and preprocess the data by handling missing values, removing outliers, and encoding categorical variables.
- **Feature Selection:** Select relevant features that may influence churn, such as account balance, transaction frequency, customer age, etc.
- **Labeling:** Define a churn event, which could be when a customer closes their account or stops using the bank's services for a certain period.
- **Data Splitting:** Split the data into training and testing sets to train and evaluate the model.
- **Model Selection:** Choose appropriate machine learning or statistical models for churn prediction. Common choices include logistic regression, decision trees, random forests, or more advanced techniques like gradient boosting or neural networks.
- **Model Training:** Train the chosen model on the training data.
- **Model Evaluation:** Evaluate the model's performance on the testing data using metrics like accuracy, precision, recall, and F1-score.
- **Hyperparameter Tuning:** Optimize the model's hyperparameters to improve its performance.
- **Interpretability:** Ensure that the model's predictions are interpretable, especially in a regulated industry like banking.
- **Deployment:** Deploy the trained model in the banking system to continuously monitor customer churn.
- **Monitoring and Maintenance:** Regularly monitor the model's performance and retrain it as needed to adapt to changing customer behavior.

Feedback Loop: Incorporate customer feedback and any new data sources to improve the model's accuracy over time.

Customer Retention Strategies: Develop strategies for retaining at-risk customers identified by the model, such as personalized offers or proactive customer support.

Compliance: Ensure that the entire process adheres to data privacy regulations and industry standards.

Remember that the effectiveness of churn prediction models can vary based on the quality of data and the chosen algorithms. It's essential to continuously refine the model and strategies to reduce customer churn and enhance customer satisfaction.

7. WORKING MODULES

7.1 Experimental Results

However, in the banking sector, customer churn prediction typically involves the use of machine learning models and historical customer data to make predictions about which customers are likely to leave or churn. The specific results would depend on the data, model, and methodology used in the experiment.

8.CONCLUSION

In an era marked by relentless technological advancements, the banking industry has witnessed a transformative shift in the way it interacts with customers and the level of service it provides. With customers having an array of options at their disposal, banks are continuously striving to not only attract new clients but, more crucially, retain existing ones. In this pursuit, the development of predictive models has emerged as an invaluable tool, allowing banks to anticipate customer churn and understand the reasons for their departure. This project has centred on Bank Customer Churn Prediction and Reason for Leaving Prediction using machine learning, with a commitment to delivering accessible and indepth insights through a user-friendly interface and data visualization. Python 3.7 has served as the foundational programming language for this project, leveraging a comprehensive suite of libraries and frameworks to accomplish the goal of customer retention prediction. The following key packages and Python libraries played a pivotal role in our approach:

❖ TensorFlow and Kera's: These deep learning frameworks are at the heart of our predictive models. They have enabled the creation of intricate neural networks to effectively model customer behaviour and churn.

❖ Pandas: The Pandas library has been instrumental in data manipulation and preprocessing. It has streamlined the handling of datasets and allowed us to structure the data for training and evaluation.

❖ Matplotlib: Data visualization is paramount for providing insights and making informed decisions.

❖ Scikit-Learn: This powerful library has been employed for machine learning tasks, including feature selection, model evaluation, and hyperparameter tuning. In conclusion, the use of machine learning in bank customer churn prediction is a pivotal step toward a customer-centric approach to banking. The implementation of Python, along with the specified libraries and frameworks, empowers banks to make informed decisions and, ultimately, cultivate lasting relationships with their customers. This project marks a step forward in the ongoing quest to make banking not only more technologically advanced but also more customer-friendly and responsive. It is a testament to the potential of machine learning in the financial sector and its ability to drive innovation and customer satisfaction

9.ACKNOWLEDGEMENT

We express our sense of gratitude towards our project guide Prof.Mr.R.Phadtare for his/her valuable guidance at every step of study of this project, also his/her contribution for the solution of every problem at each stage. We are thankful to Dr. S.M.Patil Head, Department of Computer Engineering all the staff members and project Coordinator Prof. who extended the preparatory steps of this project. We are very much thankful to respected Principal Dr. M. S. Rohakale for his support and providing all facilities for project. Finally, we want to thank to all our friends for their support & suggestions. Last but not the least we want to express thanks to our family for giving us support and confidence at each and every stage of this project.

10. REFERENCE

[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 froma database perspective,” IEEE Transactions on Knowledge and dataEngineering, vol. 8, no. 6, pp. 866–883, 1996.

[5].Y. Tang, “Deep learning using linear support vector machines,” arXivpreprint arXiv:1306.0239, 2013.

[6]. 365datascience.com/tutorials/python-tutorials/how-to-build-a-customerchurn-predictionmodel-in-python 8:

[7]. www.kaggle.com/code/kmalit/bank-customer-churn-prediction