CUSTOMER CHURN PREDICTION

1 B. Sai Vivek, 2 K. Abhishek Kumar, 3 T. Harsha Vardhan   
*Department of Computer Science and Engineering,*  
*Amrita School of Computing,*  
*Amrita Vishwa Vidyapeetham,*   
Bengaluru,Karnataka.  
1 [bl.en.u4aie22107@bl.students.amrita.edu](mailto:bl.en.u4aie22107@bl.students.amrita.edu) , 2 [bl.en.u4aie22120@bl.students.amrita.edu](mailto:bl.en.u4aie22120@bl.students.amrita.edu) , 3 [bl.en.u4aie22159@bl.students.amrita.edu](mailto:bl.en.u4aie22159@bl.students.amrita.edu)

*Abstract*— This Customer Churn Prediction is a vital strategy for businesses facing the challenges of customer attentions in today’s highly competitive market. By using the latest and advanced machine learning algorithms like Logistic Regression, Random Forest and more companies can accurately forecast customer churn, enabling them to address customer intention and attrition, optimize resource allocations, improve the customer satisfaction, it also increases the company’s revenue, and lead the company to the profits and gain a competitive advantage. This approach allows businesses to identify customers at risk of leaving, implement targeted retention strategies, and reduce the cost of hiring the new customers, ultimately fostering loyalty and long-term success in a great-evolving business landscape.

# INTRODUCTION

Churning refers to the process in which a customer leaves one company and switches to another. This not only results in a loss of income but also has negative implications for overall operations, particularly in terms of Customer Relationship Management (CRM). Establishing long-term relationships with customers is crucial for institutions as they aim to expand their customer base. Service providers face challenges related to customer behavior and their evolving expectations. The present generation, which is generally more educated than previous ones, has higher demands for connectivity, innovation, and diverse policy options. This advanced knowledge has led to changes in consumer purchasing behavior, presenting a significant challenge for service providers to think creatively and meet these expectations.

Customers may easily move their relationships from one bank to another. Some customers may keep their relationship status null, which signifies their account status is inactive. By leaving this account dormant, the consumer may be moving their connection to another bank. There are several categories of consumers in the bank. Farmers are one of the banks' most important customers; they may expect lower monthly charges because their income is modest. Businesspeople are also essential consumers since they do a large number of transactions with large sums of money. These consumers will anticipate higher levels of service excellence. Middle-class clients were one of the most significant segments; in almost every bank, these people outnumber other types of customers. These individuals will anticipate lower monthly fees, improved service quality, and new policies. Keeping multiple sorts of clients is therefore difficult. They must consider clients and their wants in order to overcome these problems and provide quality service on time and within budget to customers. Maintaining a strong working relationship with them is also a huge problem for them. If they do not overcome these eight difficulties, they may have churn. Recruiting a new client is more expensive and difficult than retaining existing consumers. Customers holding, on the other hand, are often more costly since they have already earned the trust and loyalty of existing customers. As a result, the requirement for a system that can successfully forecast client attrition in the early phases is critical for any banking institution.

# PURPOSE

Churn prediction refers to the process of identifying customers who are likely to leave a company and switch to another. The purpose of using churn prediction is to proactively address customer attrition and reduce the negative impact it has on a company's operations, particularly in terms of Customer Relationship Management (CRM). The ability to establish and maintain long-term relationships with customers is crucial for businesses aiming to expand their customer base. In today's highly competitive market, customers have higher demands for connectivity, innovation, and diverse policy options. This, coupled with changes in consumer purchasing behavior, presents significant challenges for service providers. Customer churn not only results in a loss of income but also entails the expense and difficulty of acquiring new customers. It is more cost effective to retain existing customers by accurately predicting their likelihood of churn and implementing targeted retention strategies.

# LITERATURE SURVEY

The paper "Customer Churn Prediction using Machine Learning: Subscription Renewal on OTT Platforms" by O.R. Devi, S.K. Pothini, M.P. Kumari, S.V., and U.N.S. Charan explores why people might stop using services like Netflix or Hulu. They tried different methods, like using various tools from a toolbox, to predict this. Their data included things like who the customers are, what they like to watch, and how they sign up. They measured their predictions, kind of like checking their score in a game, using metrics like accuracy and recall. But they had some challenges, like using old data and not explaining how they got it ready. They suggest that in the future, researchers could try fancier methods and look at things like social media to make predictions better. [1]. In "Machine Learning Approaches" by R. Srinivasan, D. Rajeswari, and G. Elangovan, the authors look into why customers might leave a telecom company. They tried different computer methods, like using different tools from a toolbox, to guess this. Their data included stuff like who the customers are, how they use services, and what they pay. But they didn't say how they got the data ready, which could affect their guesses. They suggest that in the future, researchers could try fancier methods and look at things like customer feelings or social media to make guesses better. [2]. The paper "Customer Churn Prediction Using Machine Learning: Commercial Bank of Ethiopia" by M.H. Seid and M.M. Woldeyohannis focuses on customer churn prediction in the context of the Commercial Bank of Ethiopia, utilizing machine learning techniques. The authors employed various algorithms, including Logistic Regression, Decision Tree, Random Forest, and Artificial Neural Networks (ANN), to develop predictive models. The study utilized a dataset obtained from the Commercial Bank of Ethiopia, consisting of customer-related information such as demographics, transaction history.[3]. In "Customer Churn Prediction Using Machine Learning" by V. Agarwal et al., the authors tackle the challenge of foreseeing when customers might leave. They tried out various computer tricks, like Logistic Regression and Decision Tree, to guess this. They measured how good their guesses were using metrics like accuracy. One problem they faced was that most customers didn't actually leave, which could make their guesses less accurate. They suggest trying different strategies to make their guesses better, like paying attention to how customers feel on social media. [4]. The paper "Customer Churn Prediction using Machine Learning" by R.K. Peddarapu et al. looks into foreseeing when customers might leave a telecom company. They tried out different computer techniques like Logistic Regression, Decision Tree, Random Forest, and Artificial Neural Networks (ANN) to make these predictions. They used data from the company and noticed there were more customers who stayed than left, which made their predictions less accurate. They suggested making predictions better by including customer feedback and considering how customer behavior changes over time. [5]. The paper "A Survey on Customer Churn Prediction in E-commerce" by P. Gopal and N.B. MohdNawi gives an overview of how companies predict when customers might leave. They looked at different studies, checking out the methods, data used, and what they measured. They used data covering stuff like who the customers are, what they bought, and how they act. They measured their guesses using things like accuracy and recall. They found problems like not having enough data, trouble picking the right things to look at, and not considering how things change over time. They suggest trying new methods and looking at things like social media and customer reviews to make better guesses in the future. [6]. In "Customer Churn Prediction Using Machine Learning Methods: A Comparative Analysis" by H. Karamollaoğlu, İ. Yücedağ, and İ.A. Doğru, the authors compare different ways computers can guess if customers will leave. They looked at methods like Logistic Regression, Decision Tree, Random Forest, SVM, and ANN. They found problems like not dealing with situations where there are more customers staying than leaving, not picking the right things to look at, and not trying out more advanced computer tricks. To make guesses better, they suggest things like using more data or picking out the important things to look at. They also think trying fancier computer tricks and looking at what people say on social media could help make guesses more accurate [7]. The paper "A Machine Learning Model for Customer Churn Prediction using Cat Boost Classifier" by J. Jane Rubel Angelina et al. is all about figuring out when customers might leave using a special computer tool called the CatBoost classifier. They made and tested a machine learning model with data from a telecom company, which included details about customers, like who they are, what they do, and how they pay. They split the data into parts for training and testing the model, and then checked how good it was using different measures like accuracy and recall. For future work, they suggest trying out new methods to make predictions even better. [8]. The paper "E-Commerce Customer Churn Prediction Scheme Based on Customer Behavior Using Machine Learning" by P. Nagaraj et al. introduces a way to predict when customers might stop shopping online, focusing on how they behave. They came up with a method that uses machine learning to make these predictions based on what customers do. They used data from an e-commerce site, including stuff like what customers buy and how they browse. They split the data into parts to train and test the model, checking how well it works using measures like accuracy and recall. While they didn't point out specific problems, they mentioned that the data might not be perfect and that there might be more of one type of customer than another. For future work, they suggest trying out fancier computer tricks like deep learning to make predictions even better. [9]. The paper "A Smote-Based Churn Prediction System Using Machine Learning Techniques" by A.O. Akinrotimi et al. is all about making a system to guess when customers might leave, using a mix of computer learning and a special balancing technique called SMOTE. They wanted to make their guesses better by dealing with situations where there are more of one kind of customer than another. They used data from a telecom company, including stuff like who the customers are and how they use services. They balanced the data using the SMOTE technique. For future work, they suggest trying new methods to make guesses even better. [10].

# PROPOSED METHODOLOGY

**Data Validation/ Cleaning/Preparing Process:**

Importing the library packages and loading the specified dataset. To investigate the variable. Identifying data by form and type, as well as analyzing missing and duplicate values. The methods and techniques for cleaning data will differ depending on the dataset. The primary goal of data cleaning is to detect and remove errors and anomalies to increase the value of data in analytics and decision making

**Exploration data analysis of visualization:**

Data visualization is an important skill in applied statistics and machine learning. Statistics does indeed focus on quantitative descriptions and estimations of data. Data visualization provides an important suite of tools for gaining a qualitative understanding algorithm or the same algorithm multiple times to form a more powerful prediction model. The random forest algorithm combines multiple algorithms of the same type i.e. multiple decision trees, resulting in a forest of trees, hence the name "Random Forest". The random forest algorithm can be used for both regression and classification tasks

**Data Visualization:**

Data visualization techniques such as histograms, scatter plots, and correlation matrices can be used to understand the distribution and relationships of the features in the datasets. This helps to identify patterns and trends in the data that can inform the development of churn prediction models.

**Model Building and Evaluation:**

The next step is to build churn prediction models using machine learning algorithms such as logistic regression, random forest, and gradient boosting. The models are evaluated using performance metrics such as accuracy, precision, recall, and F1-score. The best model is selected based on its predictive performance for churn prediction.

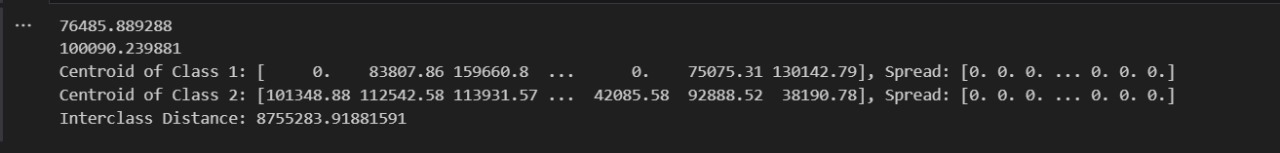
**Implementation and Action:**

The churn prediction model is integrated into the banking companies' systems to proactively identify and address potential customer churn. Customer retention strategies are developed and implemented based on the churn prediction insights. This can include targeted marketing campaigns, personalized offers, and improved customer service.

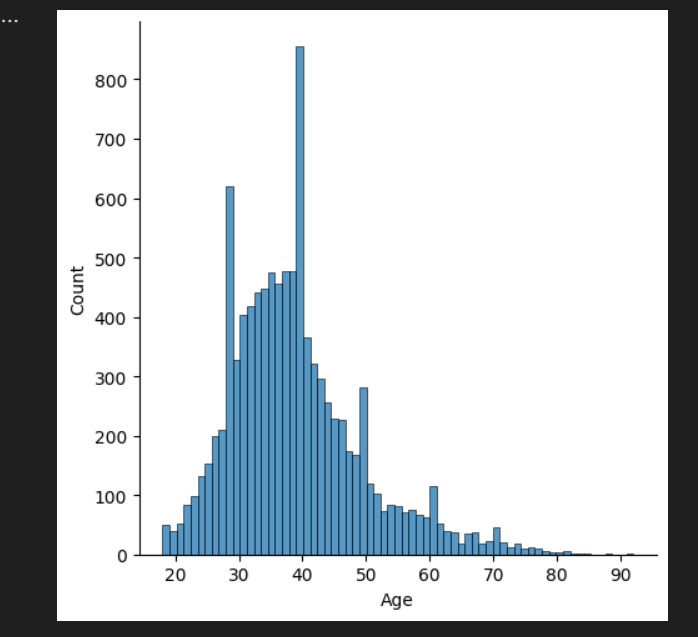
**Continuous Monitoring and Improvement:**

The performance of the churn prediction model is regularly monitored and updated as needed to adapt to changes in customer behavior and market dynamics. This involves ongoing data collection, preprocessing, and model building to ensure the model remains accurate and effective in predicting and preventing customer churn.

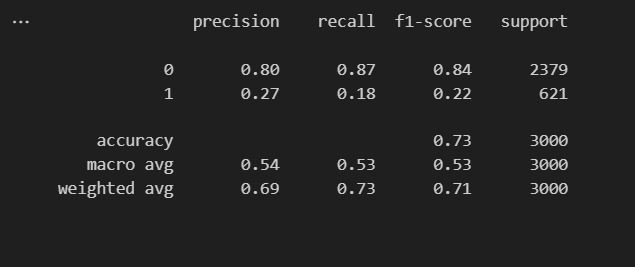
# RESULT



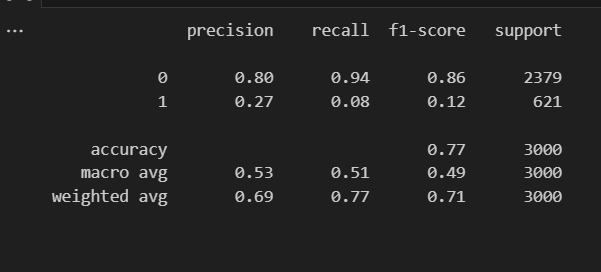
*FIG1: Mean, variance, inter class distance, centroid difference*



*FIG2: Histogram graph*



*FIG3: Implementation of KNN with K=1*



*FIG4: Implementation of KNN with K=4*

*And out data is Overfitting*

# CONCLUSION

In this project, we explored the performance of various machine learning algorithms, including Random Forest, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Grid Search, in predicting customer churn. The results showed that Random Forest achieved the highest accuracy of 85.6%, demonstrating its robustness in capturing complex relationships and interactions between variables. While KNN and SVM also exhibited competitive accuracy rates of 83% and 85% respectively, the selection of the best algorithm should consider factors beyond accuracy alone, such as interpretability, computational complexity, and scalability, tailored to the specific needs and characteristics of the business. Additionally, the lower accuracy of Grid Search emphasizes the importance of further exploration and fine-tuning of hyperparameters to enhance the predictive power of the model. Overall, the findings highlight the superiority of Random Forest in this specific project, but the algorithm choice should be made based on a comprehensive assessment of various factors in the given business context.

# REFERENCES

[1] O. R. Devi, S. K. Pothini, M. P. Kumari, S. V and U. N. S. Charan, "Customer Churn Prediction using Machine Learning: Subcription Renewal on OTT Platforms," 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2023.

[2] R. Srinivasan, D. Rajeswari and G. Elangovan, "Customer Churn Prediction Using Machine Learning Approaches," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023.

[3] M. H. Seid and M. M. Woldeyohannis, "Customer Churn Prediction Using Machine Learning: Commercial Bank of Ethiopia," 2022 International Conference on Information and Communication Technology for Development for Africa (ICT4DA), Bahir Dar, Ethiopia, 2022.

[4] V. Agarwal, S. Taware, S. A. Yadav, D. Gangodkar, A. Rao and V. K. Srivastav, "Customer - Churn Prediction Using Machine Learning," 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2022.

[5]R. K. Peddarapu, S. Ameena, S. Yashaswini, N. Shreshta and M. PurnaSahithi, "Customer Churn Prediction using Machine Learning," 2022 6th International Conference on Electronics, Communication and Aerospace Technology, Coimbatore, India, 2022.

[6] P. Gopal and N. B. MohdNawi, "A Survey on Customer Churn Prediction using Machine Learning and data mining Techniques in E-commerce," 2021 IEEE Asia- Pacific Conference on Computer Science and Data Engineering (CSDE), Brisbane, Australia, 2021.

[7] H. Karamollaoğlu, İ. Yücedağ and İ. A. Doğru, "Customer Churn Prediction Using Machine Learning Methods: A Comparative Analysis," 2021 6th International Conference on Computer Science and Engineering (UBMK), Ankara, Turkey, 2021.

[8] J. Jane Rubel Angelina, S. J. Subhashini, S. Harish Baba, P. Dheeraj Kumar Reddy, P.V. Sudheer Kumar Reddy and K. Sameer Khan, "A Machine Learning Model for Customer Churn Prediction using CatBoost Classifier," 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2023.

[9] P. Nagaraj, V. Muneeswaran, A. Dharanidharan, M. Aakash, K. Balananthanan and C. Rajkumar, "E-Commerce Customer Churn Prediction Scheme Based on Customer Behaviour Using Machine Learning," 2023 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2023.

[10] A. O. Akinrotimi, R. O. Ogundokun, M. A. Mabayoje, R. A. Oyekunle and M. O. Adebiyi, "A Smote-Based Churn Prediction System Using Machine Learning Techniques," 2023 International Conference on Science, Engineering and Business for Sustainable Development Goals (SEB-SDG), Omu-Aran, Nigeria, 2023.