University of Essex
Department of Mathematical Sciences

MA981: DISSERTATION

MACHINE LEARNING AND DEEP LEARNING-BASED CUSTOMER ATTRITION PREDICTION IN BANKING

Muhammad Jawad Arshad
2202635

Supervisor: Guo, Wenxing

December 15, 2023
Colchester

Acknowledgement

I want to express my gratitude to Dr. Wenxing Guo for his help in the dissertation. I would like to thank him for his constructive feedback and support in completing this dissertation. I want to appreciate his availability through this journey. It is essential to thank for his support in the completion of the dissertation. I also would like to thank my family, especially my parents, brother, Dr Junaid and my sister, Dr Javeria. In the end, I want to mention the name of my friend, Yue Guo, who helped me a lot in this process

Abstract

Customer attrition, sometimes called customer churn, is how customers break up their relationship with a business or organization. In the banking sector, predicting customer attrition is crucial because it allows banks to lower expenses associated with obtaining new customers, enhance client retention tactics, and better serve their customers' demands. Banks can minimize attrition rates and enhance the overall customer experience by implementing tailored retention tactics based on identifying clients who are prone to attrition. Financial institutions can successfully minimize customer attrition by customizing their offerings and utilizing focused strategies by gaining a thorough grasp of the many elements that lead to client turnover. The primary objective of this project is to do an exploratory data analysis (EDA) and develop a model to predict churn capable of reliably predicting client attrition. This will be done by using prior data and adding various causes. The principal goal of this study is to offer the bank comprehensive analysis and suggestions on customer attrition, enhancing customer loyalty, and building lasting relationships with clients.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Statement of Purpose

Nowadays, the market is known for its intense competition and constant change. This is because there are a wide range of service providers available. The goals of the present consumer generation are loftier than those of previous generations. They also have different needs regarding personalized approaches, new technologies, and increased connectivity. When viewed alongside the experiences of previous generations of customers, this is a dramatic change. They have a solid educational foundation and a thorough understanding of the strategy-making process. Due to their increasing access to information, consumers often suffer from "analysis paralysis," in which they overanalyze purchasing decisions and sales. Ultimately, this helps individuals make better selections regarding the products they buy.

The next generation of service providers is responsible for thinking outside the box to satisfy customer needs and improve customer relations in light of this enormous challenge. Identifying and respecting the client base is important for achieving success in the enterprise. The rising competitive constraints on organizations necessitate the development of novel marketing tactics to match consumer expectations, boost client loyalty and retention, and address the demands of a highly competitive market. In today's highly competitive market, the strategy of catering to everyone with more offerings is no longer effective. Instead, there is a need for a focused agenda that maximizes the effectiveness of marketing resources. This is necessary due to the growing competitiveness of the market environment. The use of technology has helped firms maintain their competitive edge over their competitors, and this pattern is expected to continue.

1.2 Introduction of Customer Churn

Data mining methods, a subset of information technology, are used in several situations, including the extraction of marketing insights and providing additional guidance for corporate decision-making[30]. This is a widely used use of data mining methodology. Since transferring is a very easy operation, customers may easily move to an organization (bank) in search of better service quality or pricing rates.

Customers have the option to transfer providers if they come across a more favorable offer or superior service elsewhere. There are a lot of businesses that are under the mistaken impression that increasing the number of new customers involves a much higher level of effort, both in terms of time and money, in comparison to maintaining connections with current customers[11]. Yet, in addition to this, the notable challenge they have is delivering reliable service to clients on time and within budgetary limitations, all while preserving a strong collaborative relationship with those customers. This is a problem that they encounter often [18]. To identify solutions for these issues, it is important to consider the demands of consumers. Furthermore, one of their main focal points will be the rate of client attrition they encounter [2]. The word "customer churn" refers to the process by which customers or subscribers withdraw their affiliation or discontinue their allegiance to a particular company or service[29]. Including as many marketing and sales personnel as possible throughout the sales funnel, the first stage for every business aiming to grow its clientele by acquiring new customers is to make full use of all available resources[9]. On the contrary, the practice of retaining existing clients usually brings greater economic benefits than acquiring new ones, owing to the organization's pre-existing relationship built upon trust and loyalty among its current clientele. This is due to the fact that the organization has established a history of successful transactions with its existing clientele. Therefore, it is important for each organization to have a system that can accurately predict the rate of client attrition during the early stages. This statement is particularly accurate during the first phases of an organization's building when it aims to develop a solid and lasting foundation.

The objective of this study is to provide a complete framework for accurately forecasting client attrition.

By using several machine learning and deep learning techniques in the banking sector, we can accurately forecast the loss of clients. Customer Relationship Management (CRM), an all-encompassing strategy, centers on the cultivation, administration, and enhancement of loyal and enduring customer relationships. It is widely acknowledged and has been actively used in several sectors, such as the retail industry, banking and insurance, and telecommunications. It is applied in a variety of industries like telecommunication, retail, banking, etc. Its main focus is to establish a relationship with the clients. Maintaining a good relationship with the existing customer is very important because acquiring a new customer is expensive and time-consuming.

Acquiring a new client (In some instances, the cost is twentyfold higher) is expensive. Hence, tools that facilitate the development and implementation of customer retention models (churn models) are essential for business intelligence (BI). These models play an important part in customer retention. Customer attrition could potentially result from inadequate levels of customer satisfaction, aggressive rival strategies, new products, regulations, and so on. It is possible that churn happened because of these factors or other factors. The purpose of churn models is to read customer behavior and decide whether this customer will churn or not. If customers will churn, it is important to take steps to stop them before the churn happens. In the past decade, there has been a notable rise in the need for research relating to various sectors, including but not limited to the gaming industry, financial sector, insurance industry, and telecommunications industry that describe the phenomena of customer churn.

1.3 Customer Retention with Machine Learning in Industries

Customer churn, alternatively referred to as customer attrition, denotes the moment when a client ceases utilizing the offerings of a particular organization's products or services. The incidence of client attrition significantly impacts the financial performance of businesses, especially in sectors that rely heavily on subscription-based revenue models.

Instances of such industries include the banking, telecommunications, internet service provider, payTV, and insurance sectors. According to approximations, the cost of acquiring a new customer could potentially be five times more expensive than the investment required to retain an existing customer.

Therefore, the study of customer attrition is of considerable significance to organizations because it empowers them to:

- Determine the weaknesses of their services, including insufficient customer support, inadequate product or service quality, or a lack of coordination with the intended audience.
- Develop strategic decisions that are well-informed and have the potential to increase customer satisfaction and, consequently, customer retention rates.

It is essential that the organization possess a precise forecast of the customer's behavior in order to find a solution to the mentioned issue. In the context of customer attrition management, two strategies are available for selection:

- A reactive approach, and
- A proactive approach

The reactive technique involves the company waiting to take any action until a cancellation request is received from the customer. Upon receiving the request, the business presents the client with appealing plan alternatives in order to maintain the client as a paying customer[2]. By adopting a proactive approach, one can foresee the potential occurrence of customer churn and subsequently present clients with the strategies they need. The current matter concerns binary classification, which involves the challenge of differentiating churners and nonchurners as separate entities. The effectiveness of machine learning in forecasting future outcomes using past data has been extensively demonstrated, which makes it a powerful tool for tackling the present situation.

This methodology utilizes a variety of machine learning algorithms, including logistic regression, support vector machines, random forest classifiers, gradient boosting classifiers, extreme gradient boosting classifiers, light gradient boosting machines, artificial neural networks, recurrent neural networks, and long short-term memories (abbreviated as ANN, RNN, and LSTM).

In machine learning models, the feature selection process becomes an essential stage in enhancing the accuracy of classification once the first preprocessing phase has been completed. To accomplish this goal, one must carefully choose the features that will be implemented in the model. Researchers have created a wide variety of approaches for feature selection.

Various strategies are beneficial in reducing the number of dimensions and computing complexity and preventing overfitting. While the churn prediction procedure is being carried out, specific attributes for forecasting churn are derived from the given information vector. This method facilitates the production of precise predictions. It is crucial for all organizations to prioritize client retention as a key objective. Since the costs associated with attracting customers usually surpass those required to maintain existing ones. The notion that clients who have established a long-standing connection will continually show loyalty can no longer be guaranteed, considering the present market situation, which offers various other possibilities. Hence, it is vital for firms to possess the ability to proactively detect consumers who are more likely to churn and implement preventive measures to mitigate this risk[9].

Predictive models pertaining to client churn possess the capability to showcase the comprehensive rate of attrition. Furthermore, understanding the fluctuation of turnover rates over different time periods, client segments, product types, and other relevant factors can yield a multitude of beneficial information. However, it is essential to recognize that customers exhibit significant variations in their behaviors and preferences. Hence, a simplistic analysis based on general assumptions or rules of thumb is unsuitable in this specific situation. Using a predictive model that utilizes gradient boosting can provide numerous advantages in this specific scenario.

1.4 Background and Justification

When compared to the process of acquiring new customers, maintaining a relationship with an existing consumer is not just more difficult but also less expensive. This is due to the fact that it is essential to allocate an appropriate budget to marketing activities in order to develop one's consumer base effectively. It is of the utmost importance to ensure that the existing client base is retained, particularly in light of the increasing number of competitors over the course of time[29].

Customers tend to leave an establishment gradually rather than abruptly. This implies adopting a proactive strategy to anticipate customer attrition by examining the past buying behaviors associated with a customer's account. Concretely, this illustrates the feasibility of forecasting client attrition[12]. Because all transactions are entered through the point-of-sale system and saved in databases, it is feasible to understand the requirements and patterns of customers as a result of the accessibility of the data[18]. This facilitates the monitoring of consumer behavior. Marketing budgets are being allocated by executives toward campaigns that prioritize the preservation of relationships with current consumers[29].Many of the numerous models that have been developed to predict client turnover employ well-known machine-learning techniques, including logistic regression and random forest. Other examples include decision trees and support vector machines[9].

This article specifically examines two separate aspects of forecasting client turnover in the banking sector. The first one is selected by the qualities that will be handed down from the model to its descendants [22]. Instead of utilizing the individuals' purchasing patterns to group them together, these values will be generated as characteristics and transmitted to the model. Subsequently, such features will be utilized. As a result, distinct features are created for each individual consumer to allow the model to gain expertise and identify unique patterns for each person [34]. Because of this, first of all, two datasets are created: one is a test data set, and the other is training data. We divide the dataset by a ratio of 80/20: 80 percent for training and 20 percent for testing. Both of these datasets will have the same data in them [18]. The second aspect that requires consideration is the implementation of the algorithm. An innovative feature of this study involves utilizing deep learning techniques to address the challenge of forecasting client attrition in the banking industry [31]. To the best of our knowledge, this is the first study to apply deep learning to work in this industry. One of the benefits of using deep learning is that it can discover prints that were earlier concealed in the given dataset[26].

Considering the fact that the cost of attracting new customers exceeds the cost of selling to present customers, it is critical for a business to track its client attrition rate. This metric assesses the degree of accomplishment an entity has attained in its endeavors.

Effective customer retention efforts lead to a rise in the client's total value over time [31], hence enhancing the value of all future sales and improving unit profits [34].In other words, a good consumer retention measure boosts the client's average lifetime deal. Instead of financing and acquiring new clients, the perfect method for a business to make the most of its current resources is to concentrate on rising earnings from trustworthy customers and recurring subscriptions on a frequent basis[9]. Utilizing available resources in the most efficient manner is the most effective approach for maximizing a business's potential. Enduring financial hardship becomes significantly more challenging when companies allocate funds toward acquiring new clients to compensate for the loss of existing ones[22]. Conversely, sustaining long-term customer loyalty significantly facilitates growth without requiring additional investment.

1.5 Defferences in The Evidence

Maintaining client satisfaction is a challenging endeavor that firms across many sectors will confront. Customers often cease their patronage of a bank due to several factors, including:

- Competitors provide comparable items at lower rates, via word-of-mouth marketing or by negative marketing activities conducted via social media.
- The competitor may have better customer service.

The organization is in the process of modifying its objectives, and the experience of its consumers is rapidly progressing to become the most essential factor. Because of this, companies need to offer a range of benefits to their customers so that their clients don't go to one of their rivals. One bad experience with a customer could be all it takes for that person to decide to start doing business with one of your competitors instead. Having a comprehensive understanding of your clients to thorough examination of their requirements and preferences is essential for operating a prosperous business.

Customer satisfaction with a company's products or services increases the likelihood of their continued loyalty to the company's brand. By utilizing statistical methodologies and machine learning algorithms, businesses may find previously unknown patterns of customer behavior and trends within the existing data [8]. Businesses can utilize this to discover unexplored possibilities for growth. Organizations can gain a competitive edge by employing data mining techniques to forecast customer attrition and improve their connection with current customers.Utilizing data mining techniques to forecast customer attrition can enable organisations to gain a competitive edge while improving the quality of their connection with current consumers. This opportunity presents itself to businesses. One of the advantages is an increase in profitability, together with a decrease in the rate of client churn.

1.6 Objective of study

The purpose of this project is: i. In order to create an accurate machine learning model for predicting customer attrition, it is important to conduct a thorough analysis. ii. Facilitating organizations in identifying consumers who show signs of possible churn and effectively implementing strategies to retain them. The main objective of this project is to get a thorough comprehension of client churn and construct prediction models within the particular context of a banking organization. Our investigation will begin with the utilization of Exploratory Data Analysis (EDA) to find and illustrate the various aspects that contribute to customer churn. With the goal of predicting customer turnover, the following analysis will offer assistance in creating machine learning and Deep learning models. In the field of classification, this issue is considered a frequent hurdle. Concerning the standards for choosing a measure of performance to improve our ML models, the job description is vague. The need to specifically categorize events associated with the positive class, namely those pertaining to customers who are prone to churn because they have greater value for the bank, inspired the recall measure.

1.7 Structur of dissertation

The structure of dissertation is that in the second chapter, we do a literature review in which we see customer churn in detail and try to find the importance of customer churn analysis. In the third part of the second chapter and at the end of the chapter, we see some machine learning classification techniques like SVM, regression analysis, and logistic analysis. In the third chapter, we discuss the methodologies that we used to complete this project. In this chapter, we describe the data set, preprocessing of the dataset, and data acquisition. At the last of this chapter, we discuss exploratory data analysis and data preprocessing. In the fourth chapter, we describe the development of machine learning and deep learning models. In the last chapter, I describe the results of our models. I put all of my code in the appendix, and in the bibliography, I give the reference to where I got help to complete this dissertation.

CHAPTER 2

Literature Review

2.1 Customer Churn

Customers will discontinue using a specific product or service. "Customer churn analysis" refers to this investigation method[17]. The term can be used in various manners, but the most straightforward interpretation is that it signifies customers no longer choosing the company due to competitive factors[24]. The goal is to recognize this situation before discontinuing the provision of the existing product or service to the consumer and then implement precautionary actions[7].

2.2 Importance of Churn Analysis

The calculations of customer churn are very important for companies that rely on subscription-based revenue, such as in the insurance[17], banking, or telecommunications industries. These industries depend on subscriptions as their main source of income [1]. Data analysis is a method used to examine data and has practical uses in several fields such as client profiling, escape analysis, and prediction[23].

Furthermore, the worth of these enterprises is strongly correlated with the magnitude of their active clientele. As a result, the number of clients has grown, resulting in a corresponding rise in customer loyalty [5]. This increase is directly linked to various business indicators, including costs, profitability [9], size, investment capacity [32], and cash flow. Furthermore, the number of clients is directly correlated with the level of loyalty exhibited by the customers.

2.3 Extensive literature review

Due to the abundance of service providers, especially banks spread over multiple global locations, the market is currently characterized as dynamic and highly competitive. This is a direct consequence of economic globalization[14]. The change in end-user behavior is among the biggest issues this economic sector is currently dealing with. The banking and finance sector is accountable for receiving funds, the execution of funds, and the provision of loans, thus making its consumers the fundamental identity of the industry[10]. The financial condition of an organization's profit is closely linked to the rate at which it retains its consumers[9]. This proportion must be minimized to the greatest extent possible. As per the Harvard Business Review, a mere five percent decrease in a company's customer base can lead to a profit increase ranging from twenty-five to eighty-five percent[16]. Hence, considering the paramount importance of clients as the primary drivers of a bank's profitability, the contemporary banking sector is founded upon five vital pillars[10].

The structure comprises five pillars: asset management, capital management, liquidity management, and risk management. The effective maximization of profits by a financial institution's management relies on focusing their efforts on the five fundamental pillars that underpin the institution[18]. Hence, the continual influx of new clients poses a significant challenge for financial institutions[15]. "Customer churn" is the phrase used to describe the financial harm experienced by a firm when it loses one or more customers to a competitor[17].To efficiently handle customer churn, it is crucial to identify the clients who are at risk of transferring their banking affiliation from one institution to another competitor[8].

Moreover, to optimize the worth of the client base, it is important to prioritize churn control to develop enduring and appropriate interactions between organizations and customers [25]. The term "churn" refers to client attrition, which can be classified into two distinct categories: voluntary and involuntary[35]. Recognizing the phenomenon known as "nonvoluntary churn" is not challenging. This occurs when a bank chooses to terminate the services it offers to its customers[15].

Conversely, distinguishing voluntary churn from involuntary churn is more challenging because it entails a consumer deliberately choosing to end their business association with a certain bank [13]. This makes it more challenging to track it down. Furthermore, it is feasible to further categorize it into two separate factions, known as accidental and intentional churn, respectively."Incidental churn" is a word used to describe a situation where a customer's circumstances change in a way that prevents them from continuing their business relationship with their bank, for example, when their financial condition undergoes a shift [14]. This makes up a small fraction of the whole amount. Deliberate customer attrition can be attributed to various factors [35], including the introduction of new technological services, enhancements in pricing, and considerations of quality [20].

Financial institutions should regularly monitor their consumers to promptly identify warning signals of customer behavior that could lead to customer loss[25]. Researchers and bank management are currently analyzing the data to identify patterns and trends. The objective is to create models that can accurately predict if a customer will close their account, also known as churn[16]. The objective of these models is to keep clients from discontinuing their banking services.

Furthermore, data play a vital role in the banking industry, where clustering techniques, such as neural network categorization based on client attributes, are suggested for discovering hidden patterns inside extensive databases[4]. One of the key areas of study in machine learning technology is the classification of customers based on whether they are categorized as churning or non-churning customers[18]. The anticipation of client attrition by utilizing extensive data is a specialized area within the discipline of marketing research[17]. Over the past few decades, numerous studies published in academic journals have led to the development of a diverse range of forecast models, including linear regression, decision trees, random forests, logistic regression, neural networks, support vector machines, and deep neural networks [15]. These models are used to provide predictions regarding future events. These models are examples used in machine learning [17].

The investigation into the number of individuals who will terminate their credit cards is not new; numerous researchers have already formulated a diverse array of models to accomplish this task [19]. While developing a predictive model, we took the unique transaction histories of individual customers. The model relies heavily on behavioral trait variables, decision factors, and spatiotemporal data. This measure was implemented to ensure the generation of accurate model results[10]. Because demographic-based factors were included in the newly constructed model, the findings demonstrated that it outperformed traditional models in terms of prediction accuracy[14].

Additionally, researchers dove headfirst into developing machine learning models that could reliably forecast client churn in the first stages of the study[16]. Experts created a model for estimating how many customers will terminate their credit card accounts [17]. The three different machine learning techniques used to achieve this were k-nearest-neighbor (KNN), linear regression, and random forest. A dataset with 10,000 data points and 14 unique characteristics was indicated by the data collecting method[24].

The findings indicate that the three factors with the most significance were the aggregate transaction amount, the number of transactions in the preceding year, and the overall revolving balance [17]. The objective of the research was to discover and predict the primary causes that caused consumers to terminate or discontinue using their accounts during the last six months [4]. The results indicate that those with strong loyalty toward the financial institution and those who borrowed larger sums were less likely to close their accounts [6]. Findings indicated that the model correctly predicted operating losses of up to 10 percent at the leading Brazilian banks[19]. Another area that was investigated for attrition prediction was retail banking[14]. Based on the findings of the research, the performance of the extra-tree classifier model demonstrated a statistically significant improvement in comparison to the performance of other models. According to the study, the most important factors in creating a forecast model were the lack of mobile banking, zero-interest personal loans, zero balance, and other online services [6]. The study results reveal that the random forest model performed noticeably better than the other models that were considered.

Additionally, many researchers have built a system capable of forecasting the behavior of a wide range of customers who have churned in various enterprises like telecommunications and e-commerce [6]. Depending on the experiment's results conducted by [10], this technique can save enterprises a considerable amount of money. "Customer churn" prediction is a technique utilized in the banking business to gain knowledge of the probability of clients transferring their banking relationship to a different bank [14]. The surge may be ascribed to several causes, such as the accessibility of cutting-edge technical innovations, low-interest rates, affordable services, and the appealing attributes associated with credit cards [22].

In order to make predictions about the customers who will terminate their credit card services, the aim of this study is to make predictions based on data about the buyers, like their ages and genders [17]. Meanwhile, scholars have been exploring methods to identify fraudulent activity on credit cards via machine learning models or the development of optimizing techniques that leverage machine learning models. Each of these concepts is now under discussion for future investigation[20].

2.4 Deep learning as learning Embedding

Today, banking service providers are confronted with intense rivalry within the industry [21]. The phenomenon known as customer churn, which describes the situation in which clients cancel their existing subscriptions to a banking service, is one of the factors that contribute to the increased level of competition. It is possible to take this behavior as an example of dissatisfaction on the part of the customer. It is possible that the banking service provider will face significant challenges as a result of client attrition if the situation is not treated with diligence [23]. When clients subscribe for shorter periods, the bank must anticipate a decrease in total revenue [27]. Banks must have a thorough understanding of customer conduct, especially about those who may cancel their membership in the future[33]. This comprehension is necessary in order to guarantee the retention of customers and the generation of subscription revenue.

There are multiple strategies to accomplish this goal, including creating a coherent artificial intelligence (AI) model. This statement is significant in the telecommunications industry of developing nations, as there is a growing focus on the progress of artificial intelligence [5]. Using vector embedding as a technique can result in substantial advantages when building an interpretable model. In order to effectively execute this technique [17], the model must have the ability to generate a comprehensive vector for each unique data point it represents [1]. The vector can acquire the desired knowledge by passing explicit instructions to the model for doing specific tasks [20]. Extrapolating conclusions from the implicit information inside the vector makes it feasible to reveal underlying knowledge within the data. It is feasible to reveal hidden knowledge in the data.

By incorporating vector embedding into deep learning models, explainability can be achieved [20], allowing these models to serve as decision support systems for a banking organization. There is a substantial correlation between customer behavior and the advice given by decision support systems, which could help the organization make decisions based on consumer behavior [19]. The model can also be employed for forecasting the probability of client churn. With its capacity as a decision support system [24], a forecasting model can present alternate viewpoints and suggest preferred decisions. Deep learning is widely considered the most favored artificial intelligence (AI) method for many applications [14]. Several businesses, including retail, mobile gaming, music streaming, and telephony, have adopted this concept as their primary approach to churn modeling [9]. Although created models are widely used, their understanding remains mysterious due to the intricate structure of deep learning, making it difficult to evaluate their resultscite[20]. The surge in popularity of deep learning in recent years may be associated with this occurrence. Throughout history, several methodologies have been developed to get a better grasp of the complex dynamics behind deep learning[14]. Vector embedding is a technique utilized in natural language processing [20]. This approach allows the model to construct a vector representation, which can be analyzed to reveal the knowledge learned throughout the deep learning training process. This is a methodology that can be utilized [3]. Vector embedding has been utilized to analyze customer turnover in the mobile gaming domain[14]. However, the vectors themselves have not been analyzed. The word2vec 20 program, which is a part of Natural Language Processing (NLP), has had a substantial impact in promoting the idea of vector embedding (2022).

2.5 ML: Classification Techniques Overview

A quick review of the algorithms that are most commonly used in customer churn is below. The research community admire the efficiency [17], and extensive use [28] of these algorithms.

2.5.1 Artifial Neural Network

ANN, often known as Artificial Neural Network, is an important technique to solve complex problems like customer churn predictions[5]. This is the most common method. There are two different types of ANN: hardware-based networks, which represent the neuron through the physical components [32], and software-based neurons, which represent the neuron through the use of computer models[33]. Both types of neural networks are examples of artificial neural networks. The example of the supervised model that is recognized is the multi-layer perceptron, which represents this kind of model[2]. The training of the algorithms involves the iteration of the backpropagation algorithm, which is known as (BPN). This algorithm is a feed-forward algorithm that uses supervised machine learning[27]. For the prediction of customer churn, neural networks have better performance than decision tree[11].The experimental results show that ANN also has better results over logistic regression[16].

2.5.2 LGBMC Algorithm for Customer Churn Prediction

LGBM stands for the light gradient-boosting machine learning classifier. LGBMC is a machine learning classifier which is specialized in handling large datasets and complex features. The main benefit of this approach is the enabled technique, which increases the performance (accuracy) of the algorithm by combining lesser learners into strong predictions. The capability of the algorithm to manage categorical data and computational efficiency makes it good for the prediction of customer churn.

To indicate the customer churn, by effectively identifying the pattern, LGBM uses the gradient boosting technique for better accuracy. The performance of the algorithm improves by its ability to handle imbalanced dataset.

2.5.3 Regression Analysis-Logistic Regression Analysis

Regression analysis is a statistical method to find the strength of the relationship between many variables in the dataset [32]. When analyzing the variable, it is common practice to model and evaluate both dependent and independent variables at the same time [17]. The use of regression analysis in customer churn prediction is constrained due to the suitability of linear regression models primarily for predicting continuous values[9]. Logistic regression, or LR, is a statistical classification model with a probabilistic approach. Moreover, it produces a binary prediction of a categorical variable [5], such as customer churn. This forecast is generated using predictor variables [7], which are based on customers' attributes.

2.5.4 Support Vector Machine for Customer Churn

Examine a problem in the domain of machine learning that addresses explicitly binary classification and encompasses n features [28]. In the n-dimensional feature space, a hyperplane exists as a separator between the training data points belonging to the two separate classes [9]. This plane can alternatively be defined as a boundary [32]. The Support Vector Machine

(SVM) seeks to find a hyperplane that effectively separates two parallel hyperplanes, maximizing the distance between them [17]. The training data points on the hyperplanes with the maximum margin are known as support vectors. That is why this algorithm is called"support vector Machine" [29]. Dot products between test data points and support vectors must be run before predictions can be made [27]. SVM may be conceptualized as a constrained quadratic optimization problem in mathematical terms [23]. An additional penalty provision has been included to address concerns about the absence of segregation [9]. The kernel function is a highly efficient approach for solving nonlinear separation between hyperplanes. A nonlinear separating hyperplane in the original feature space is converted into a separating linear hyperplane in a higher-dimensional space by this function [7].

SVM is a strong instrument for identifying customer churn in many industries.SVM can create a clear difference between customers that can churn or not by using its classification techniques. It draws a line that separates both churned and non-churned customers and this is the primary goal of SVM.

CHAPTER 3

Methodology

3.1 Overview of Dataset

The dataset we are using consists of 10000 rows and 14 columns, i.e., row number, customerId, surname, CreditScore, Geography, Gender, Age, Tenure, Balence, NumberOfproduct, HasCrcard, IsActiveMember, EstimatedSalarly, Exited. Features like RowNumnber, CustomerId, and Surname are specific for every customer, so we can drop them.

Variable	Data Type
CreditScore	int (64)
Geography	Object
Gender	Object
Age	int(64)
Tenure	int(64)
Balance	float(64)
NumberOfProdcts	int(64)
HasCard	int(64)
IsActiveMember	int64
EstimatedSalarly	float(64)
Exited	int(64)

Table 3.1: Summary of Data Set Attributes and their Types

3.1.1 Feature Summary

The data set has the following characteristics, and their corresponding data types are shown in table 3.1.

CreditScore: This is the credit score of the persons.

Age: Its represents that how old is our customer.

Tenure: It shows that how long the customer belong to the bank

Balence: It shows the balance of the account.

No.OfProduct It is the representation that how many product a customer have

HasCrCard This column shows that the customer has credit card or not.

IsActiveMember: It is the representation that how many customer is acitve member.

Estimated Salarly it shows the estimated salary of customer.

3.2 Acquisition and pre-processing of the data set

3.2.1 Data Acquisition

The first phase Our techniques include the process of loading a data set. We get the dataset from Kaggle, titled as`Customer_Churn.csv`.

3.2.2 Handling Missing Values

The dataset we are using is consist of 10000 rows and 14 columns and fortunately there is no missing values in our dataset. The absence of missing values within the dataset offers several advantages to this study. Firstly, it contributes significantly to the reliability of the findings, as the entirety of the data was available for analysis. This completeness allowed for a comprehensive exploration of relationships between variables, minimizing the potential for bias or misinterpretation resulting from missing data. While the absence of missing values in the dataset contributes positively to the study and reliability, it is crucial to acknowledge its implications and consider avenues for further exploration in research methodologies.

3.3 Exploratory Data Analysis (EDA)

Expolator data analysis is very important part of the EDA. It is the first step to understand the data, and our model development almost depends on the EDA.EDA is a very important tool for researchers and analysts to deeply and thoroughly analyze the data. If we make any mistake in this step, it means that our latter steps, like model feature selection and model development will also have defects.

3.3.1 How many customers Retained or Churned

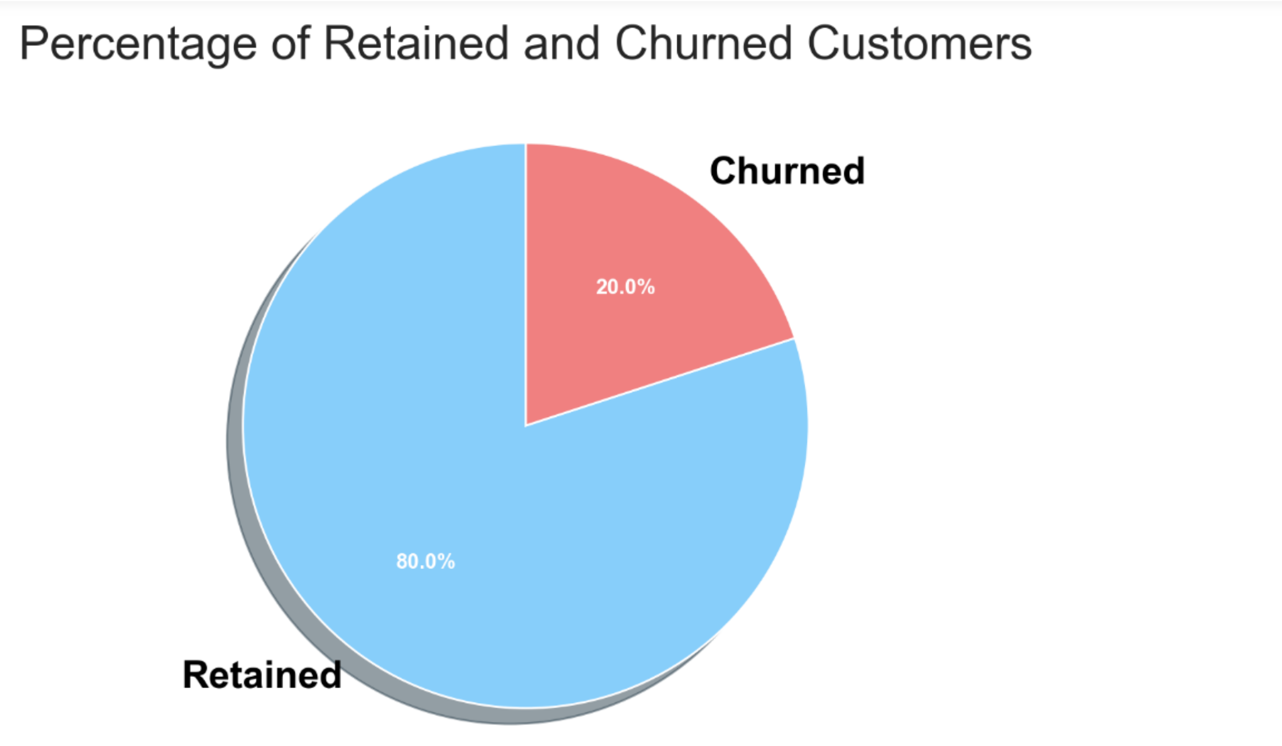

Figure 3.1: Percentage of retained customer churn

From the above picture, we assume that 80 percent of customers were retained and 20 percent left. If we talk about the number, approximately 80 percent of customers were retained, and 20 percent of customers left the bank.

3.3.2 Credit Score

From the credit score chart, primarily, the customer has a good credit score, i.e., 600, so it means that they have a good credit history, which shows that they have good financial health.

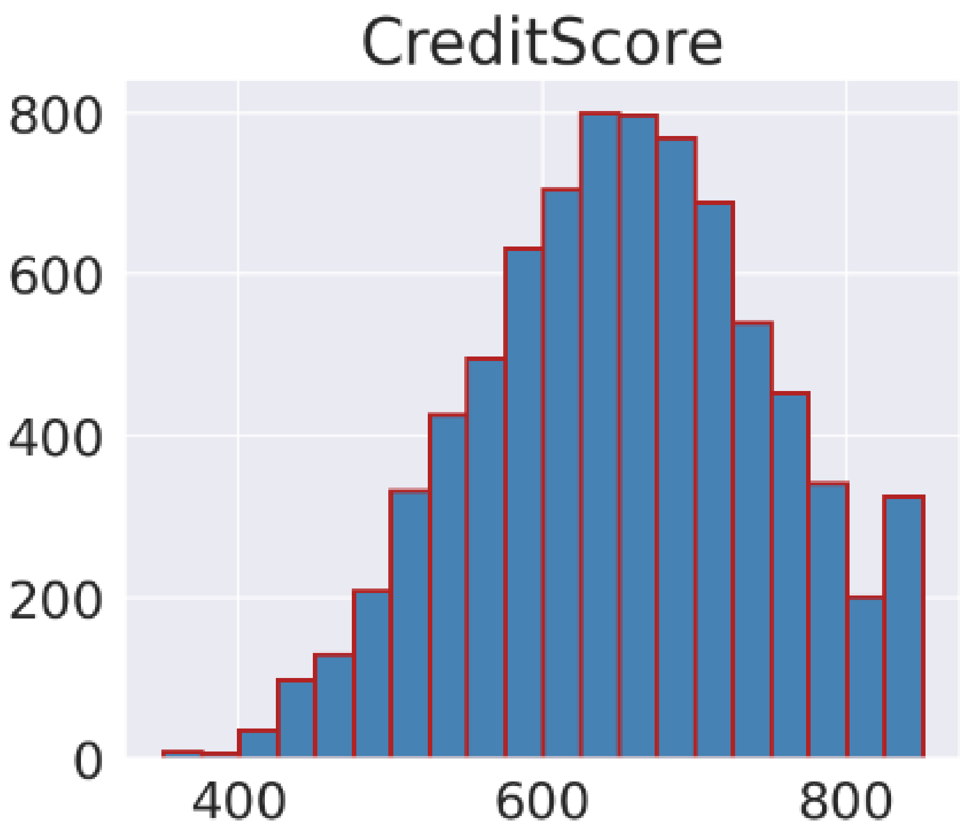

Figure 3.2: Credit Score

3.3.3 Looking for corelation

A correlation is a statistical analysis in which we see how strongly the two variables are related. Its value varies from -1 to 1. If the correlation is 1, the variables have a high positive relationship. On the other hand, if the correlation is -1, it means that the relationship is very negative, and if our correlation becomes 0, it demonstrates no relationship between variables. The formula for the correlation coefficient (r) between two variables X and Y is:

$$r = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}} \tag{3.3.1}$$

where:

- n is the number of data points,
- $\sum XY$ is the sum of the products of corresponding values of X and Y,
- $\sum X$ and $\sum Y$ are the sums of values of X and Y respectively,
- $\sum X^2$ and $\sum Y^2$ are the sums of squares of values of X and Y respectively.

3.3.4 Corelation of continuous Features

The figure shows that there is no intercorrelation (intercorrelation means the relation between the features within the dataset) between the features because the dark blue color shows a strong relationship between the features. However, from the figure, we visualize no box close to the dark color, so we do not need to worry about the multicollinearity.

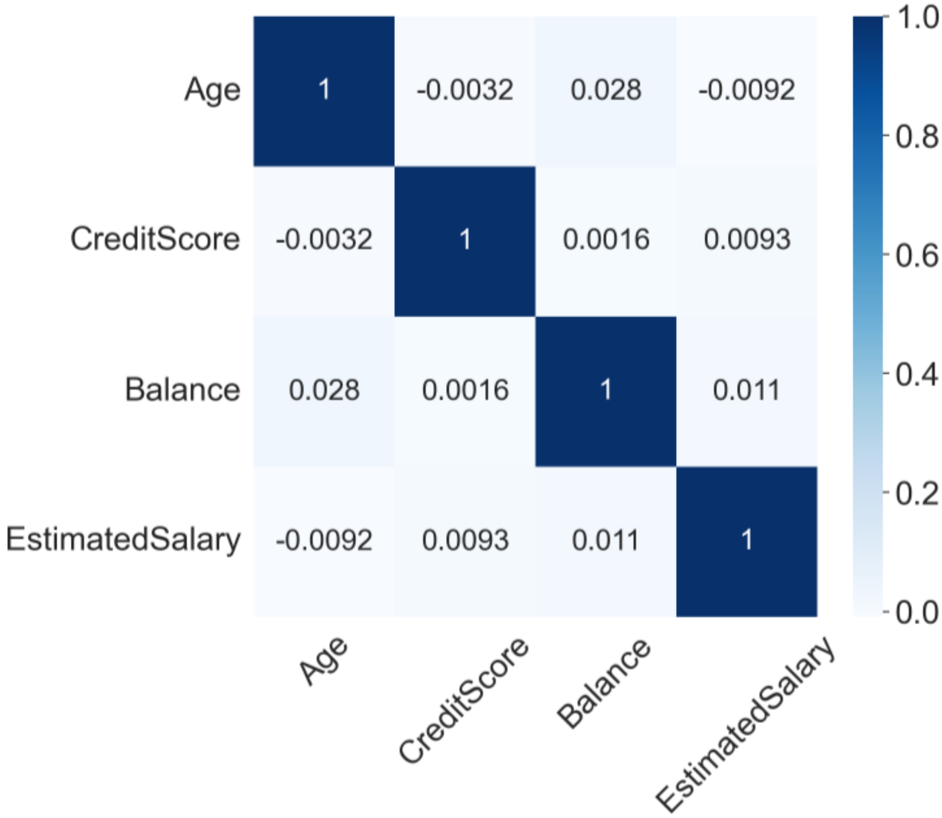

Figure 3.3: Corelation of continuous variable

Now, we look into each feature separately in detail.

3.3.5 Age

From the figure below, we know older people have more chances to churn. Maybe it is because the bank is not able to meet the needs of the older people, and this figure also has another thought that the older people do not want or do not need to use banks anymore because of their limited needs. From the figure, we also know that after age 50, the customers start to churn.

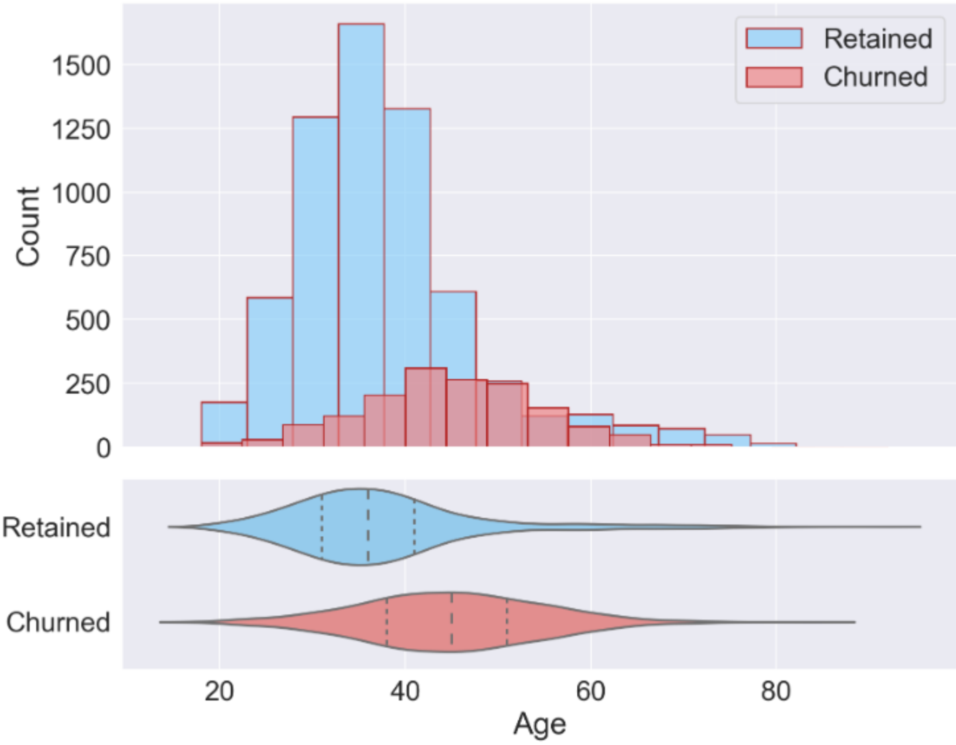

Figure 3.4: Expected age of customer churn

3.3.6 Crdit Score

After looking deeply into the dataset, it is clear that there is no difference between the credit scores of the customers who retained or churned. So, from this, it is clear that churn is not the only feature on which we decide whether the customer will retain or churn.

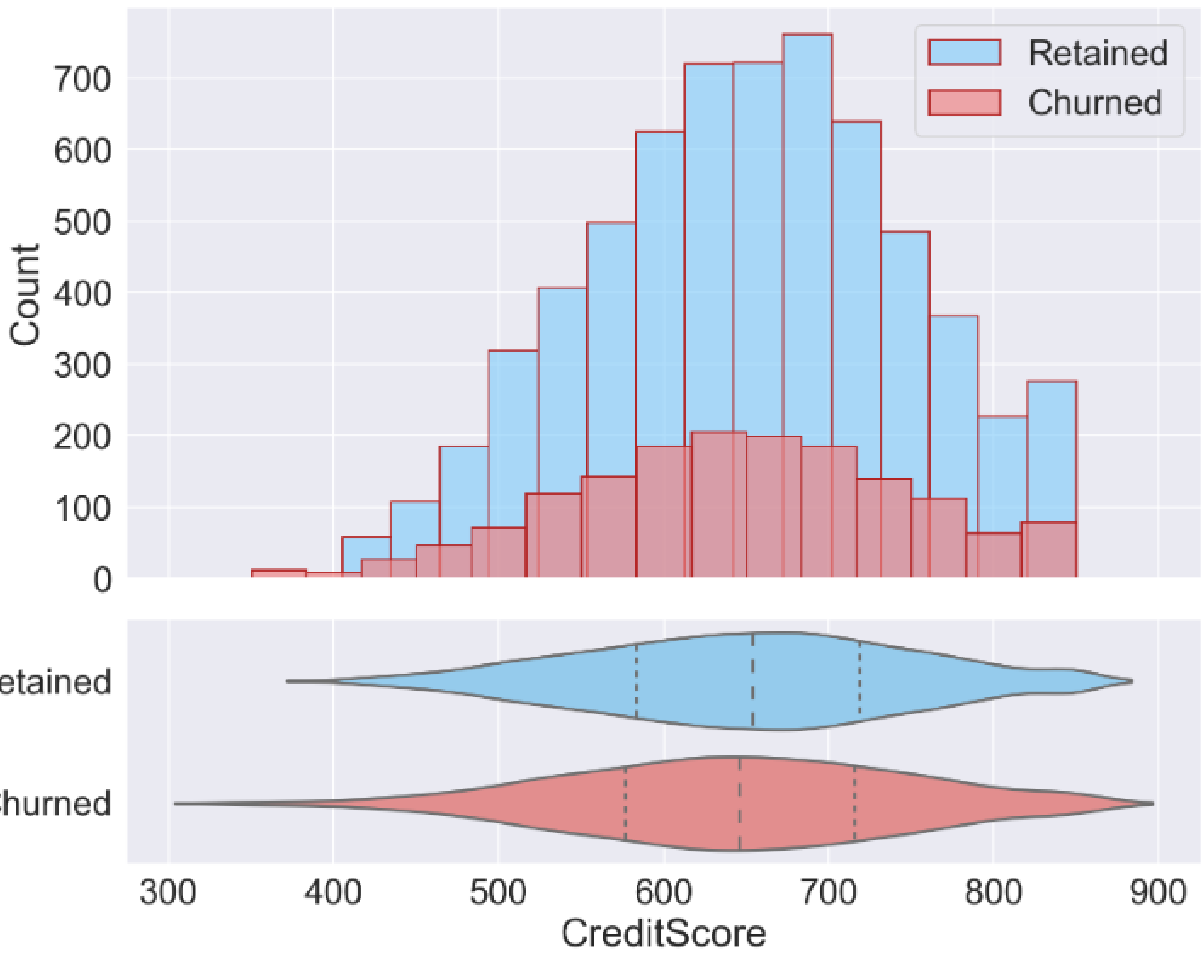

Figure 3.5: Credit Score of Expected Churned customer

3.3.7 Balance

The figures show that customers with high balances will likely be churn.

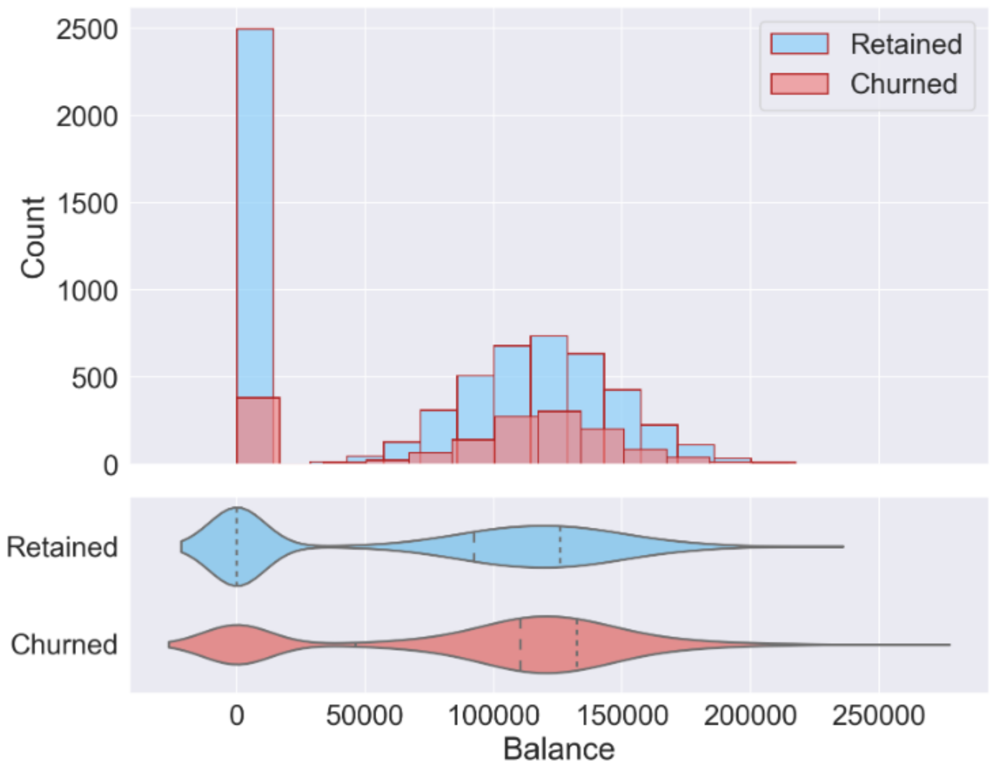

Figure 3.6: Balence of Expected of Customer Chunr

3.3.8 Categorical Variables

Categorical variables are variables that represent a group of data or categories. This type of variable does not have numbers. We use categorical variables when we need to analyze the groups. From the figure below, we can extract the following points.

- The bank has customers from three countries: Germany, Spain, and France. Most customers are French.
- The number of male customers is greater than that of females, so we can say that mainly bank audiences consist of males.
- A tiny percentage of customers leave the bank. The tenure of the customer from years 1 to 10 is almost the same
- Most customers buy two products from the bank. A few customers buy three products, and approximately 0 buy four.
- Nearly 80 percent of customers have credit cards.

- Roughly 4,000 customers are not active.

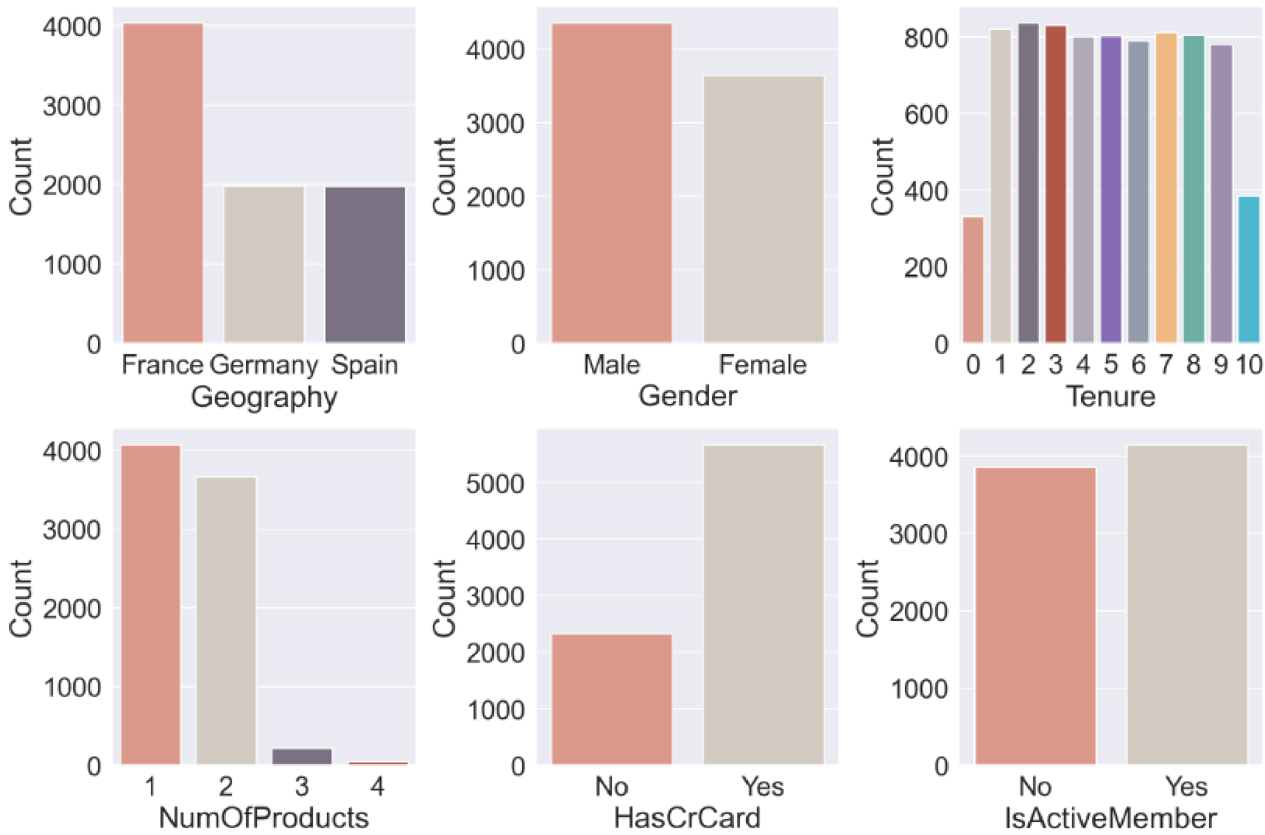

Figure 3.7: Categorical Features of dataset

3.4 Data Pre-Processing

Data preprocessing transforms raw data into a format that can be easily used. It includes many phases: cleaning, transformation, feature engineering, normalization, feature selection, categorical feature encoding, scaling, and correcting class imbalance.

However, in our project, we use the following strategies for our data preprocessing:

- Feature selection.
- Encoding categorical feature.
- Scaling.
- Adressing Class imbalance.

Now we apply each technique separately in our project one by one.

3.4.1 Feature Selection

As in Section 3.1, we already dropped some columns like RowNumber, CustomerId, and Surname. We will further dig into our dataset by applying the chi-square test and finding p values.

3.4.2 Chi-square test

The chi-square test is a type of statistical method in which we find whether the features/variable of the dataset has some relationship between them or not. In easy words, we can use this test to find whether the features depend on data or not. This type of test is widely used in health, biology, mathematics, market research, social sciences, data science, etc.

The formula for the chi-square test statistic (χ^2) is:

$$\chi^2 = \sum \frac{(O-E)^2}{E} \tag{3.4.1}$$

Where:

$$\begin{aligned} \chi^2 &: \text{Chi-square test statistic} \\ O &: \text{Observed frequency in a cell} \\ E &: \text{Expected frequency in a cell} \\ \sum &: \text{Summation symbol over all cells in the contingency table} \end{aligned}$$

3.4.3 p Value

The p-value is a statistical metric used to assess the level of evidence against the null hypothesis in hypothesis testing. A number indicates the probability that data is observed or whether the null hypothesis is true.

The formula for the p-value in a hypothesis test is:

$$\text{p-value} = P(\text{observed results} \mid \text{null hypothesis is true} \tag{3.4.2}$$

Where:

$$
\begin{aligned}
\text{p-value} &: \text{Probability of observing results} \\
&\quad \text{as extreme as or more extreme than the observed results} \\
&\quad \text{if the null hypothesis is true} \\
P &: \text{Probability}
\end{aligned}
$$

3.4.4 Applying Sci Square and p-Value

Variable	Chi-square	p-value
NumOfProducts	1233.595	3.767e-267
Geography	230.748	7.829e-51
IsActiveMember	195.315	2.199e-44
Gender	90.173	2.183e-21
Tenure	15.197	1.250e-01
HasCrCard	0.301	5.833e-01

Table 3.2: Chi-square Test Results

The variables 'Tenure' and 'HasCrCard' have a low chi-square value and a p-value greater than 0.05 (the commonly used threshold), confirming our original premise that these attributes do not provide substantial information. Therefore, we choose to eliminate these variables.

3.4.5 Encoding Categorical Features

Machine learning and deep learning models require that their input and output variables will be in numerical form. Therefore, we need to convert categorical features into numbers before building ML and DL models, this process is called encoding features. So, for that reason, we take the following steps.

- For the feature Gender, we encode male as 1 and female as 0.

- The bank's audience is from 3 countries: Germany, France and Spain. The customer churn in Germany is more significant than in the other two countries, so under this observation, we deal with Spain and France as the same and Germany separately. So, we encode Germany as 1 and France and Spain as 0

3.4.6 Scaling

Scaling, in the context of data analysis and machine learning refers to the process of altering the range of a distribution or numerical variable. Some of the few types of scaling are as follow:

- Normalization.
- Standardization.
- MinMax Scaling.

However, in this project, we use the standardization technique. In data analysis, standardization is usually used to change the numerical value so that specific properties have a mean of 0 and a standard deviation of 1. The formula for standardization (Z-score normalization) is:

$$\text{Standardized value (Z-score)} = \frac{x - \mu}{\sigma} \tag{3.4.3}$$

Where:

$$\begin{aligned} x &: \text{A data point in the dataset} \\ \mu &: \text{Mean (average) of the dataset} \\ \sigma &: \text{Standard deviation of the dataset} \\ \text{Standardized value (Z-score)} &: \text{Transformed value after standardization} \end{aligned}$$

So, after applying standardization, we successfully split our dataset into X train and y train sets.

3.4.7 Class Imbalance

Class imbalance is a classification problem when the dataset's distribution of classes (categories or labels) is highly skewed, meaning one class has significantly more or fewer instances than the others. This problem comes in many real-world senarios where one class is more prominent than the other. Some solutions may be used to address this problem. For this project, the SMOTE (Synthetic Minority Oversampling Technique) technique is used to address this issue.

CHAPTER 4

Model development

4.1 Logistic Regression

We aim to categorize our observations into two categories when we analyze our data, which shows a binary divide: consumers who are likely to stop using the bank and customers who are unlikely to stop using it. Finding out how likely it is that a person belongs to a particular group is what the logistic regression model is all about. The projected result is limited to a predefined range in logistic regression, a generalized version of linear regression. The model aims to use probabilistic calculations to classify customers into appropriate groups by establishing relationships between the target characteristic, Churn, and other parameters.

The logistic regression hypothesis is defined as:

$$h_\theta(x) = \frac{1}{1 + e^{-(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_n x_n)}} \tag{4.1.1}$$

where:

$$\begin{aligned} h_\theta(x) &: \text{Logistic regression hypothesis} \\ \theta_0, \theta_1, \ldots, \theta_n &: \text{Model parameters (weights)} \\ x_1, x_2, \ldots, x_n &: \text{Input features} \\ e &: \text{Base of the natural logarithm (Euler's number)} \end{aligned}$$

The logistic function $\frac{1}{1+e^{-z}}$ is also known as the sigmoid function, and it maps any real-valued number z to the range $(0, 1)$. The logistic regression model predicts the probability that the dependent variable is 1 given the values of the independent variables.

4.2 Support Vector Classifier

In the current competitive corporate environment, the expectation of client attrition has emerged as a major concern. The accuracy rate, an important statistic for businesses, often falls short of expectations even though churn prediction has been the subject of much study. The strong basis of Support Vector Machines (SVMs) in statistical learning theory has led to their popularity in many disciplines, including data mining, computer vision, pattern recognition, and machine learning[29]. This may primarily be defined by their exceptional accuracy and resilient capacity to formulate extensive generalizations. Machine learning makes heavy use of Support Vector Machines (SVM), a technique based on statistical learning theory. The presented method effectively addresses complex challenges such as nonlinearity, high dimensionality, and localized reduction in forecasting client churn for the banking sector. Frequently, conventional methods prove inadequate in adequately tackling these difficulties. A practical method for predicting customer turnover entails using a Support Vector Machine (SVM) in conjunction with structural risk minimization.

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \tag{4.2.1}$$

Subject to the constraints:

$$\begin{cases} \mathbf{w} \cdot \mathbf{x}_i + b \geq 1 & \text{if } y_i = 1 \\ \mathbf{w} \cdot \mathbf{x}_i + b \leq -1 & \text{if } y_i = -1 \end{cases}$$

Here:

$\mathbf{w}$ is the weight vector,

$\mathbf{x}_i$ is the input feature vector,

b is the bias term,

y_i is the class label for the i-th data point.

Also included is a new metric for evaluating models, which considers covering, lift coefficient, and hit rate. This study's technique borrows heavily from earlier work that presented Support Vector Machines (SVM) for this purpose. After that, the evaluation criterion was reviewed with the overall accuracy rate in consideration. To further address and decrease erroneous decisions, weight SVM was also used. We can ensure an optimum separating hyperplane and classification decision function if we can effectively separate the customer data along a hyperplane near the data vector plane. This is the underlying premise of the proposed technique. If the data vectors are linear and inseparable, then a new penalty for sample misclassification during separation must be included [29]

4.3 Random Forest

Random Forest Regression is a sophisticated predictive model that uses decision trees to make predictions. These decision trees resemble flowcharts since they evaluate data points at each decision node according to specific requirements. During the training phase, the algorithm identifies many decision points and builds trees based on the variable values from the training dataset. For numerical data, predictions are either made by calculating the average values of the data points in the leaf nodes or by assigning predicted categories based on the categories of the data points in the leaf nodes. The Random Forest ensemble constructs several trees, usually about 100, and trains them using a bootstrap aggregation technique. The accuracy of forecasts may be affected by the number of trees and their maximum depth. Random Forest classification or regression techniques are non-parametric and can effectively handle skewed and categorical data, whether ordinal or non-ordinal. Renowned for their precision, these models are adept at efficiently managing a substantial number of autonomous predictor variables. Data preparation starts with thoroughly comprehending the data that is accessible for analysis. This includes both categorical data fields, such as marital status and gender, and continuous numeric data fields, such as tenure and age.

In a decision tree, the Gini impurity (tell us about the probability of misclassifying an observation.) at node t for a binary classification problem is given by:

$$G(t) = 1 - \left(p_{\text{class\_0}}^2 + p_{\text{class\_1}}^2\right) \tag{4.3.1}$$

Where:

$G(t)$ is the Gini impurity at node t,

$p_{\text{class\_0}}$ and $p_{\text{class\_1}}$ are the probabilities of belonging to class 0 and class 1, respectively.

The decision tree algorithm uses this impurity measure to select the best split at each node based on features, aiming to minimize impurity.

A decision tree is used to identify customer churn in the banking sector because it has the ability to find an important factor that causes churn, identify the change in the pattern of customer behavior, and has the capacity to work efficiently with large datasets. It also helps to group the customer based on their behavior so that from this, bank management comes to know which customer will likely churn or which customer will not churn.

4.4 Gradient Boosting Classifier

Hyperparameter consist to of many parameter that are also part of this agorithem. How these parameters affect the model's overall performance and the depth of the underlying trees during each boosting cycle, determine how well gradient boosting can forecast customer attrition in the banking industry[23]. Performing a thorough examination of all potential options to identify the most beneficial values may be challenging and time-consuming.

Let $F_m(x)$ represent the additive model at iteration m:

$$F_m(x) = F_{m-1}(x) + \gamma h_m(x) \tag{4.4.2}$$

Where:

F$_{m-1}(x)$ is the previous model,

$h_m(x)$ is a new weak learner (decision tree),

γ is the learning rate controlling the contribution of each tree.

Gradient Boosting updates the model in a way that minimizes the gradients of the loss function in an effort to minimize a loss function $L(y, F(x))$. The approach fits a new weak learner to the loss function's negative gradient at each iteration m:

$$h_m(x) = \arg\min_h \sum_{i=1}^{n} L\left(y_i, F_{m-1}(x_i) + h(x_i)\right) \tag{4.4.3}$$

Each new model is added to the ensemble in turn, with the primary focus being placed on the residuals or errors that are the consequence of the combined predictions of the prior models. Through the process of repeatedly optimizing the loss function, gradient boosting is able to construct a strong predictive model.

4.5 Xtreme Gradient Boosting Classifier

A common challenge faced by organizations across all sectors is the issue of accurately predicting customer attrition using massive amounts of time-series data. Because of the inherent temporal complexity of time-series data, generating model features for training and testing on multiple time periods is a challenging process. A consumer database with many temporal characteristics is analysed using extreme gradient boosting (XGBoost) in this work. The primary objective of this research is to create an accurate customer-churn model. This research aims to give a solid solution to the time-sensitivity issue in feature engineering. By incrementally adding additional models, XGBoost minimizes an objective function:

$$\text{Objective Function} = \sum_{i=1}^{n} L(y_i, F(x_i)) + \sum_{k=1}^{K} \Omega(f_k) \tag{4.5.1}$$

Where:

$L(y_i, F(x_i))$ is the loss function measuring the difference between predicted and actual values,

$\Omega(f_k)$ is the regularization term penalizing model complexity,

n is the number of data points,

K is the number of weak learners (trees),

y_i represents actual target values,

$F(x_i)$ represents predicted values,

f_k represents individual weak learners.

By optimizing the complexity penalties and model predictions repeatedly, XGBoost seeks to minimize this objective function.

4.6 Light Gradient Boosting Machine

This research proposes a suggested approach for evaluating customer attrition, with the aim of enhancing sales effectiveness. This study aims to identify the factors that contribute to customer churn and develop strategies to retain more loyal customers. Customer turnover forecasting may be accomplished by the usage of machine learning technologies, specifically by using LightGBM decision tree models.

$$\text{Objective Function} = \sum_{i=1}^{n} L(y_i, F(x_i)) + \sum_{k=1}^{K} \Omega(f_k) \tag{4.6.1}$$

Where:

$L(y_i, F(x_i))$ is the loss function measuring the difference between predicted and actual values,

$\Omega(f_k)$ is the regularization term penalizing model complexity,

n is the number of data points,

K is the number of weak learners (trees),

y_i represents actual target values,

$F(x_i)$ represents predicted values,

f_k represents individual weak learners (trees).

4.7 Deep learning model development

The neurons in the human brain serve as inspiration for deep learning, a subfield of machine learning that focuses on neural networks. Using multi-layer neural networks is what the word "Deep" refers to.

4.7.1 Artificial Neural Network

An artificial neural network is a deep learning model that works like the functionality of the human brain, which consists of neurons. It is a very important component of deep learning that is usually used in pattern recognition, classification problems, regression, etc. Important components of the neural network are as follows

- Neurons:
  It is a very basic element of the neural network. They receive input, apply weights, and perform calculations.
- Layers:
  A neural network consists of multiple layers, and layers consist of multiple neurons. Usually, there are three types of layers the input layer, output layer, and hidden layer.
- Connections:
  Neurons are interconnected through the edges that transmit information from one layer to another.

Because of the hidden layers of the neuron, they are known for their ability to learn complex patterns and recognize complex patterns from the data.

Due of their capacity to comprehend complicated patterns in large datasets, banks utilise Artificial Neural Networks (ANNs) to anticipate client attrition. ANNs may detect minor churn signs by analysing transaction history, demographics, and behavioural tendencies. Advanced modelling in their multi-layered architecture helps banks estimate churn, predict client behaviour changes, and customise retention efforts to increase customer loyalty and minimise attrition. ANNs learn to detect churning clients via iterative learning, allowing banks to proactively solve retention issues.

4.7.2 Recurrent Neural Network

A recurrent neural network is a type of neural network that is built to handle sequential data. Not like feedback neural networks (output only in one direction), RNNs have the ability to form loops, allowing them to retain and use information from previous steps. The main components of RNN are as follows:

- Recurrent Connection:
  RNN has recurrent connection in which each neuron in the network receives input not only from the current step but also from its own output at the previous step.

- Hidden state:
  This algorithm has the ability to have hidden states that store data from the previous state.

- Sequence Processing:
  RNN sufficiently manages sequential types of data of many different kinds, like text, images, and time series.

The mathematical formula of RNN is as follows

$$h_t = \sigma(W_{hh}h_{t-1} + W_{xh}x_t + b_h) \tag{4.7.1}$$

- h_t is the hidden state at time t.
- σ is the activation function.
- W_{hh} and W_{xh} are the weight matrices.
- x_t is the input at time t.

Sequential data processing via Recurrent Neural Networks (RNNs) helps banks retain customers. RNNs use time-series data to predict client behavior by analyzing transaction sequences and engagement patterns. Banks may dynamically personalize retention efforts using RNNs to recognize changing client preferences and behaviors by capturing temporal relationships. Their ability to recognize and forecast sequential trends allows banks to take preemptive steps like targeted marketing or customized service offers to strengthen client relationships and reduce turnover.

4.7.3 Long Short Term Memory

LSTM is the latest version of RNN which helps to solve gradient problems and find dependencies in the sequential data.Its development helps to solve the problems of RNN by storing data over long sequence.The main features of LSTM is as follow:

- Memory Cell:
  LSTM contains a specific type of cells called memory cells which stores information over a long period. These cells consists of many gates to control the information.

- Gates:

  This algorithm consists of many different gates which decide which information has to keep and which information has to delete. The gates are as follows:

  - Input gate:

    Decide which new information is going to be stored in the cells

  - output Gate:

    It determines the output of information based on the current cell state.

  - Forget Gate:

    This gate decides which information we have to keep or which information we have to discard.

- Long Term Memory:

  Through its gating or check mechanism, the LSTM network can selectively remember or forget information over many times steps enabling it to capture long-term dependencies in sequential data

Chapter 5

Results

After all of the deep learning and machine learning models have been applied to our dataset, the next step is to examine the outcomes of our algorithms. After all, we will see the results of machine learning, and after that, we will discuss the deep learning models. Before moving to results, I would like to explain some of te terms like accuracy, precision, and recall.

- Accuracy: In machine learning and deep learning, accuracy is the term that refers to the measure of current prediction by a model over the number of total prediction models.
- Precision: It is the indicator of the machine learning model and deep learning models that measure the model performance by the quality of positive predictions made by the models.
- Recall is the matrix used in machine learning and deep learning to evaluate the performance of a model. It helps to measure the model's capacity to correctly identify all relevant instances.

5.1 Comparison Of Machine Learning Algorithms

Model	**Accuracy**	**Precision**	**Recall**
LR	0.683	0.667	0.730
SVC	0.797	0.808	0.780
RF	0.796	0.813	0.770
GBC	0.788	0.808	0.755
XGB	0.793	0.800	0.781
LGBMC	0.805	0.822	0.780
SVot	0.804	0.816	0.785

Table 5.1: Performance Metrics Of Machine Learning

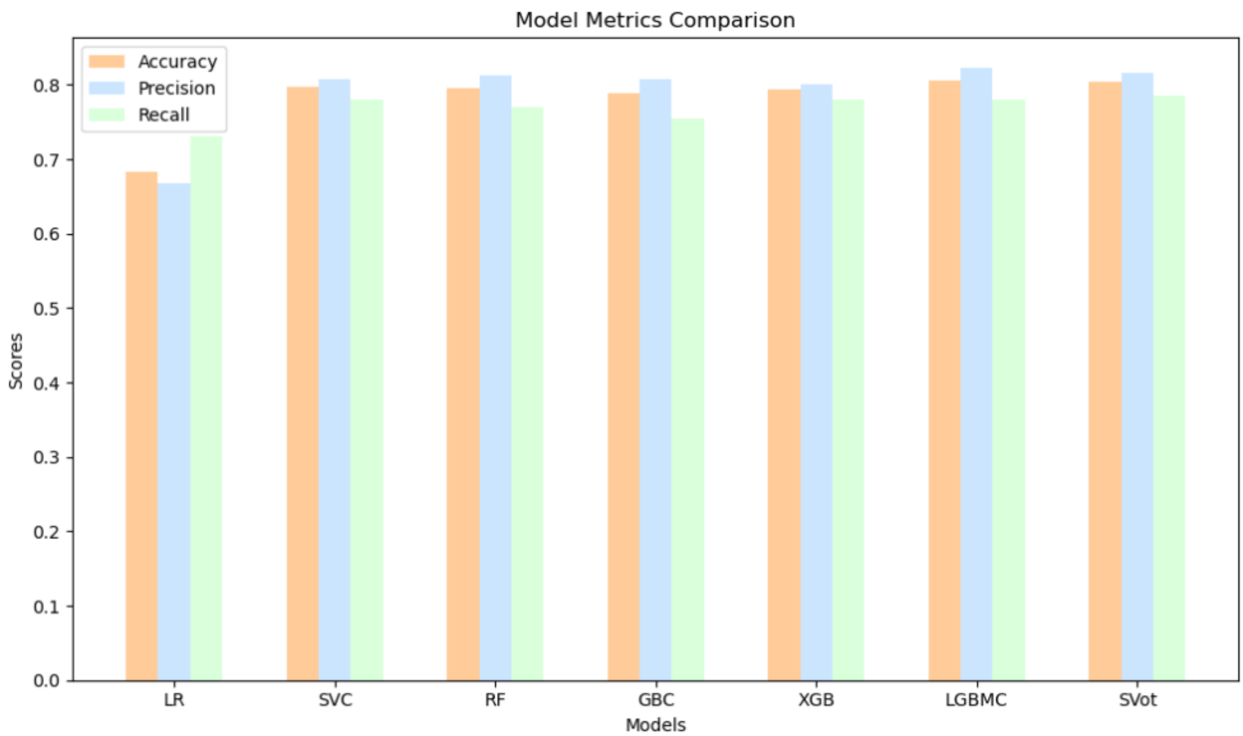

Figure 5.1: Graphical Comparison of ML Algorithms

From the table and bar chart above, we come to know that all of our algorithms have recalled, i.e., more than 70 percent. Only one algorithm, named XGB has a higher recall of 78 percent. But from an overall perspective, LGBM has the best performance, with higher accuracy and precision. So from this point of view, we can say that XGB is the best machine-learning algorithm to predict customer churn.

5.2 Comparison Of Deep Learning Algorithms

After discussing deep learning models, we will discuss deep model results.

Model	**Accuracy**	**Precision**	**Recall**
ANN	0.85	0.87	0.95
RNN	0.85	0.99	0.91
LSTM	0.86	0.87	0.97

Table 5.2: Deep Learning Model Metrics

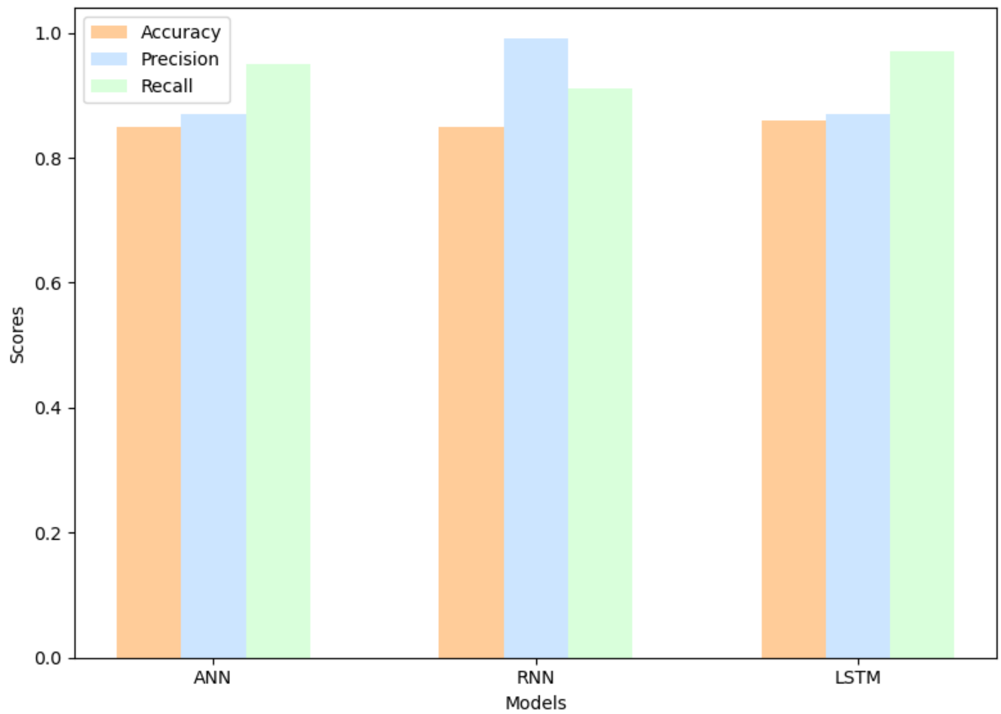

Figure 5.2: Graphical Comparison of DL algorithms comparison

So, from the above bar chart and table, we find that LSTM is our best deep learning model because we find that LSTSM is the best algorithm because it has high recall value and accuracy to predict customer churn, with the highest accuracy and higher recall.

5.3 Comparison of Deep Learning and Machine Learning

From the machine learning algorithms, we know that XGB is the best model with an accuracy of 0.805, precision of 0.822, and recall of 0.785, but from the deep learning models, LSTM is the best algorithm. So the comparison of machine learning and deep learning algorithms, LSTM is the best overall algorithm to predict the customer churn in the banking sector.

APPENDIX A

Appendix

Here is the link to the source code that is already uploaded on GitHub

https://github.com/jawadarshad123/Final-Desertation.git

APPENDIX B

Data Set and Resources Utilized

B.1 Data Set Utilization

The following URL contains the dataset that was utilised in this study:

https://www.kaggle.com/code/korfanakis/predicting-customer-churn-with-machine-learning/inp

B.2 Resources Utilized

The report was formatted in Latex using the online Latex compiler Overleaf, while the project was implemented in Python 3.10.11 using the renowned Jupyter Notebook compiler.

https://github.com/jawadarshad123/Final-Desertation.git

Bibliography

[1] Essam Abou el Kassem, Shereen Ali Hussein, Alaa Mostafa Abdelrahman, and Fahad Kamal Alsheref. Customer churn prediction model and identifying features to increase customer retention based on user generated content. *International Journal of Advanced Computer Science and Applications*, 11(5), 2020.

[2] Sanket Agrawal, Aditya Das, Amit Gaikwad, and Sudhir Dhage. Customer churn prediction modelling based on behavioural patterns analysis using deep learning. In *2018 International conference on smart computing and electronic enterprise (ICSCEE)*, pages 1–6. IEEE, 2018.

[3] Samer M Arqawi, MA Abu Rumman, EMAN AKEF Zitawi, et al. Predicting employee attrition and performance using deep learning. *J Theor Appl Inf Technol*, 100, 2022.

[4] Pushkar Bhuse, Aayushi Gandhi, Parth Meswani, Riya Muni, and Neha Katre. Machine learning based telecom-customer churn prediction. In *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, pages 1297–1301. IEEE, 2020.

[5] Shulin Cao, Wei Liu, Yuxing Chen, and Xiaoyan Zhu. Deep learning based customer churn analysis. In *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*, pages 1–6. IEEE, 2019.

[6] Ozer Çelik and Usame O Osmanoglu. Comparing to techniques used in customer churn analysis. *Journal of Multidisciplinary Developments*, 4(1):30–38, 2019.

[7] Tjeng Wawan Cenggoro, Raditya Ayu Wirastari, Edy Rudianto, Mochamad Ilham Mohadi, Dinne Ratj, and Bens Pardamean. Deep learning as a vector embedding model for customer churn. *Procedia Computer Science*, 179:624–631, 2021.

[8] Alae Chouiekh et al. Deep convolutional neural networks for customer churn prediction analysis. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, 14(1):1–16, 2020.

[9] Anouar Dalli. Impact of hyperparameters on deep learning model for customer churn prediction in telecommunication sector. *Mathematical Problems in Engineering*, 2022:1–11, 2022.

[10] Edvaldo Domingos, Blessing Ojeme, and Olawande Daramola. Experimental analysis of hyperparameters for deep learning-based churn prediction in the banking sector. *Computation*, 9(3):34, 2021.

[11] Juan Pablo Equihua, Henrik Nordmark, Maged Ali, and Berthold Lausen. Modelling customer churn for the retail industry in a deep learning based sequential framework. *arXiv preprint arXiv:2304.00575*, 2023.

[12] Juan Pablo Equihua, Henrik Nordmark, Maged Ali, and Berthold Lausen. Modelling customer churn for the retail industry in a deep learning based sequential framework. *arXiv preprint arXiv:2304.00575*, 2023.

[13] Samah Wael Fujo, Suresh Subramanian, Moaiad Ahmad Khder, et al. Customer churn prediction in telecommunication industry using deep learning. *Information Sciences Letters*, 11(1):24, 2022.

[14] Priya Gopal and Nazri Bin MohdNawi. A survey on customer churn prediction using machine learning and data mining techniques in e-commerce. In *2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, pages 1–8. IEEE, 2021.

[15] Yunxuan He, Ying Xiong, and Yiting Tsai. Machine learning based approaches to predict customer churn for an insurance company. In *2020 Systems and Information Engineering Design Symposium (SIEDS)*, pages 1–6. IEEE, 2020.

[16] Pankaj Hooda and Pooja Mittal. An optimized kernel msvm machine learning-based model for churn analysis. *International Journal of Advanced Computer Science and Applications*, 13(5), 2022.

[17] Nagaraju Jajam, Nagendra Panini Challa, Kamepalli SL Prasanna, and Venkata Sasi Deepthi Ch. Arithmetic optimization with ensemble deep learning sblstm-rnn-igsa model for customer churn prediction. *IEEE Access*, 2023.

[18] Sanjay Kumar and Manish Kumar. Predicting customer churn using artificial neural network. In *Engineering Applications of Neural Networks: 20th International Conference, EANN 2019, Xersonisos, Crete, Greece, May 24-26, 2019, Proceedings 20*, pages 299–306. Springer, 2019.

[19] Praveen Lalwani, Manas Kumar Mishra, Jasroop Singh Chadha, and Pratyush Sethi. Customer churn prediction system: a machine learning approach. *Computing*, pages 1–24, 2022.

[20] Rencheng Liu, Saqib Ali, Syed Fakhar Bilal, Zareen Sakhawat, Azhar Imran, Abdullah Almuhaimeed, Abdulkareem Alzahrani, and Guangmin Sun. An intelligent hybrid scheme for customer churn prediction integrating clustering and classification algorithms. *Applied Sciences*, 12(18):9355, 2022.

[21] Mehul Mahrishi, Kamal Kant Hiran, Gaurav Meena, and Paawan Sharma. *Machine learning and deep learning in real-time applications*. IGI global, 2020.

[22] Abinash Mishra and U Srinivasulu Reddy. A novel approach for churn prediction using deep learning. In *2017 IEEE international conference on computational intelligence and computing research (ICCIC)*, pages 1–4. IEEE, 2017.

[23] Saifil Momin, Tanuj Bohra, and Purva Raut. Prediction of customer churn using machine learning. In *EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing: BDCC 2018*, pages 203–212. Springer, 2020.

[24] Jajam Nagaraju and J Vijaya. Methodologies used for customer churn detection in customer relationship management. In *2021 international conference on technological advancements and innovations (ICTAI)*, pages 333–339. IEEE, 2021.

[25] Maciej Pondel, Maciej Wuczyński, Wiesława Gryncewicz, Łukasz Łysik, Marcin Hernes, Artur Rot, and Agata Kozina. Deep learning for customer churn prediction in e-commerce decision support. In *Business Information Systems*, pages 3–12, 2021.

[26] R Prashanth, K Deepak, and Amit Kumar Meher. High accuracy predictive modelling for customer churn prediction in telecom industry. In *Machine Learning and Data Mining in Pattern Recognition: 13th International Conference, MLDM 2017, New York, NY, USA, July 15-20, 2017, Proceedings 13*, pages 391–402. Springer, 2017.

[27] Manas Rahman and V Kumar. Machine learning based customer churn prediction in banking. In *2020 4th international conference on electronics, communication and aerospace technology (ICECA)*, pages 1196–1201. IEEE, 2020.

[28] Aryan Raj and D Vetrithangam. Machine learning and deep learning technique used in customer churn prediction:-a review. In *2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)*, pages 139–144. IEEE, 2023.

[29] Akshara Santharam and Siva Bala Krishnan. Survey on customer churn prediction techniques. *International Research Journal of Engineering and Technology*, 5(11):3, 2018.

[30] A Saran Kumar and D Chandrakala. A survey on customer churn prediction using machine learning techniques. *International Journal of Computer Applications*, 975:8887, 2016.

[31] Philip Spanoudes and Thomson Nguyen. Deep learning in customer churn prediction: unsupervised feature learning on abstract company independent feature vectors. *arXiv preprint arXiv:1703.03869*, 2017.

[32] R Sudharsan and EN Ganesh. A swish rnn based customer churn prediction for the telecom industry with a novel feature selection strategy. *Connection Science*, 34(1):1855–1876, 2022.

[33] Hoang Tran, Ngoc Le, and Van-Ho Nguyen. Customer churn prediction in the banking sector using machine learning-based classification models. *Interdisciplinary Journal of Information, Knowledge & Management*, 18, 2023.

[34] V Umayaparvathi and K Iyakutti. Automated feature selection and churn prediction using deep learning models. *International Research Journal of Engineering and Technology (IRJET)*, 4(3):1846–1854, 2017.

[35] Wanli Xing and Dongping Du. Dropout prediction in moocs: Using deep learning for personalized intervention. *Journal of Educational Computing Research*, 57(3):547–570, 2019.