Customer Churn Prediction with Machine Learning Approaches

BY

Prattoy Paul Borson
ID: 191-15-2540

AND

Anika Tahsin
ID: 191-15-2658

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Al Amin Biswas
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

Md. Sabab Zulfiker
Senior Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY DHAKA, BANGLADESH

FEBRUARY 2023

APPROVAL

This Project/internship titled **"Customer Churn Prediction with Machine Learning Approaches"**, submitted by Prattoy Paul Borson and Anika Tahsin, ID No: 191-15-2540 and 191-15-2658 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 01 February 2023.

BOARD OF EXAMINERS

Chairman

Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Tania Khatun (TK)
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Ms. Lamia Rukhsara (LR)
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

Dr. Mohammad Shorif Uddin
Professor
Department of Computer Science and Engineering
Jahangirnagar University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Al Amin Biswas, Senior Lecturer, Department of CSE,** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma. This presentation has been held on *1 February, 2023*.

Supervised by:

29.1.2023

Al Amin Biswas
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

29.01.23

Md. Sabab Zulfiker
Senior Lecturer
Department of CSE
Daffodil International University

Submitted by:

Prattoy Paul Borson

Prattoy Paul Borson
ID: 191-15-2540
Department of CSE
Daffodil International University

Anika Tahsin

Anika Tahsin
ID: 191-15-2658
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes it possible for us to complete the final year project/internship successfully.

We are really grateful and wish our profound indebtedness to **Al Amin Biswas, Senior Lecturer,** Department of CSE, Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Machine Learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to **Al Amin Biswas, Md. Sabab Zulfiker,** and **Professor Dr. Touhid Bhuiyan,** Head**,** Department of CSE, for his kind help to finish our project and also to other faculty members and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Customer churn prediction is a critical task for many industries, such as telecommunications, banking, and e-commerce. This paper presents a comprehensive survey of customer churn prediction methods, which are typically classified into three categories: statistical methods, machine learning-based methods, and deep learning-based methods. The survey focuses on each category, introducing the most relevant approaches of churn prediction, as well as their respective strengths and weaknesses. We also discuss the challenges and open research issues related to this field. Finally, we outline the future research trends in customer churn prediction in order to inspire new research ideas. RandomForestClassifier achieved the highest accuracy of 84.00%, outperforming other machine learning and Deep learning algorithms.

Keyword: Customer, Churn, MLP classifier, SVM, KNN, Decision tree classifier, Random Forest classifier, Gaussian NB, Deep neural network.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

Chapter 1
Introduction

1.1 Introduction:

Customer churn prediction is the process of using data-driven methods to predict when a customer is likely to abandon a product or service. It is a common metric used by companies to measure customer loyalty and satisfaction. Customer churn can be caused by a variety of factors, including poor customer service, lack of product features, changes in pricing, and competitive pressure. Companies track customer churn in order to identify areas of improvement, respond to customer feedback, and develop strategies to retain customers.Customer churn prediction is a process of predicting which customers are likely to leave a company. It is a key metric for measuring the effectiveness of a company's customer service and retention efforts. By predicting which customers are likely to churn, companies can proactively take steps to reduce the risk of losing them. This may involve offering special promotions, improving customer service, or providing additional incentives to keep them engaged. Predictive models can help identify which customers are at risk of churning and offer tailored solutions to retain them.The total customer experience largely determines brand perception and has an impact on how customers assess the value of the goods or services they use. The truth is that if a brand has given them a problem or multiple problems, even devoted customers won't tolerate it.

1.2 Motivation:

I choose this subject because, based on a customer's past behavior, customer analytics enables an organization to discover both opportunities and challenges. It is turning those signals into useful information about their clients. Additionally, analytics enables the direct dissemination of customer data to individuals or organizations in a format that is easy to understand and act upon. Customer analytics will assist businesses in examining consumer behavior and attitudes around the goods and services they provide. Customer analytics can assist businesses in deciphering the customer's activities, whether they are good or poor, making it simpler for businesses and individuals to comprehend and grasp this information, resulting in better decisions being made and the success of the business being maintained.I personally want to pursue a career in customer

analytics and operations analytics because having access to customer data, such as where they shop, what they buy, when they buy, and how they pay, will enable me or my employer to predict how customers will behave under different conditions in the future by understanding their preferences and acting accordingly.

1.3 Rationale of the Study:

The study of customer churn prediction is important because it can help businesses better understand why customers leave and how to prevent it. By predicting customer churn, businesses can make adjustments to their product offerings, marketing strategies, and customer service to reduce the rate of churn and retain more customers. Additionally, understanding customer churn can help businesses better understand customer behavior and identify opportunities for growth.

1.4 Research Questions:

1. What are the most common factors that lead to customer churn?
2. How can we identify customers at risk of churning?
3. What strategies can be employed to reduce customer churn?
4. How can we best use customer data to predict churn?
5. How can customer churn be prevented?
6. What are the differences between customers who churn and those who don't?
7. How can machine learning be used to predict customer churn?
8. What types of customer segmentation can be used to better predict customer churn?
9. Can customer churn be predicted before it actually happens?
10. How can customer service teams use customer churn prediction to improve customer loyalty?

1.5 Report Layout:

1. Introduction: Provide an overview of the customer churn prediction process, including the purpose and goal.

2. Data Preparation: Describe the data preparation process, including any data cleaning or feature engineering.
3. Modeling: Explain the modeling techniques used and discuss the results of the model.
4. Evaluation: Present the evaluation metrics used to measure the accuracy of the model.
5. Conclusion: Summarize the findings and discuss any potential implications.

Chapter 2
Background

2.1 Related Works:

Service providers have identified the importance of maintaining the customers present in such a rapid setting. To increase precision and performance, the issue of customer turnover can be addressed by merging Support Vector Machines with boosting algorithms.
[1].

A novel approach has been constructed to anticipate customer churn in the telecommunications field. A support vector machine (SVM) was revealed to have the greatest precision rate, hit rate, covering rate, and lift coefficient. Equally, SVM can be utilized in the banking sector with analogous customer churn data. [2].

This research aims to identify the optimal model for anticipating customer churn in a banking institution. Different feature selection procedures lead to different outcomes with the Random Forest classifier combined with oversampling producing the highest rate of success at 95.74%. [3].

The results demonstrate that random forests perform better than decision trees, k-nearest neighbors, logistic regression and other machine learning algorithms. Our research highlights the value of investing in cross-selling and upselling strategies for current customers. This section details the data used and how it was pre-processed.[4].

A Brazilian e-commerce outlet is being studied in order to forecast customer attrition. The likelihood of customers to leave is determined by the amount from their initial purchase, the types of items purchased, and the delivery cost. Neither the population density in the customer's area nor whether it is rural or urban has an influence.[5].

The most essential part of creating a supervised model is model validation. An appropriate data division strategy must be employed to generate a model with outstanding generalizability. The proposed model achieved the top performance using 250 neurons, one hidden layer, a batch size of 32, and 4000 epochs.[6].

The project was completed with Python on Google Colab. A variety of classifiers, including LR, DT, KNN, and RF, were tested with basic features. The DT classifier had the best recall value of .4429 and the highest accuracy at 85.20%, outperforming the other classifiers.[7].

Churn prediction can be beneficial for many firms in terms of generating profits and bringing in revenue. Churn is an expression used to refer to the losses incurred when customers leave for other services or due to network issues. Machine learning approaches can be used to recognize which customers are most likely to end their subscription.[8].

Annually, Markel Corporation experiences a decrease in premiums due to clients choosing not to renew their policies. Extra trees classifier achieved the highest result of 0.68 Auc score in PLC4, while GBM scored 0.61 Auc in PLC123.[9].

In this paper, the writer can concentrate on different machine learning approaches for foreseeing customer attrition. The creator can build up categorization models, for example, Logistic Regression, Support Vector Machines, Random Forest, and Gradient Boosted Trees. The information accumulated incorporated 7043 perceptions over 21 factors in a .csv design. [10].

This research seeks to forecast client attrition in influencer commerce. Influencers sell things directly on social media after promoting them through SNS since influencer commerce is a type of online shopping. Influencers play a promotional role for businesses and/or goods on social media platforms like Twitter, Facebook, and Instagram.[11].

This research, in contrast to other studies, aims to make use of PyCaret Toolkit, an open source machine learning package created to make carrying out standard tasks in a machine learning project straightforward. Consequently, a client cluster with a greater fraud risk has been found .[12].

In order to decrease customer turnover and develop strategies to keep customers, firms will benefit from understanding the causes of customer churn and the actual customer churn rate utilizing the information provided in this paper.[14]

2.2 Complex Engineering

Customer churn prediction is a complex engineering problem that requires the use of various techniques to identify patterns in customer behavior. These techniques can include advanced statistical modeling, machine learning, artificial intelligence, and deep learning algorithms. The goal is to detect customer churn before it is too late and to provide proactive solutions to reduce the likelihood of churning. The data used for customer churn prediction is typically collected from customer interactions with the company, such as website visits, customer support interactions, customer feedback surveys, and loyalty program points. This data is then used to develop forecasts that can recognize customers who have a high chance of leaving.This allows companies to take proactive steps to retain customers and prevent them from leaving.

Chapter 3
Research Methodology

3.1 Work Process

The model for churn prediction in the proposed system will be built using Python programming.Python is an incredibly powerful tool for statisticians and data miners, allowing them to develop statistical software and analyze data. It has not yet been tapped into its full potential, particularly in terms of creating models to predict customer churn. In this work, we looked at a variety of machine learning algorithms to identify customer churn. To accurately pinpoint those customers that are likely to leave, multiple models were tested.These models are MLP Classifier, SVM, KNN, Decision Tree Classifier, Random Forest Classifier, GaussianNB,Deep neural network.

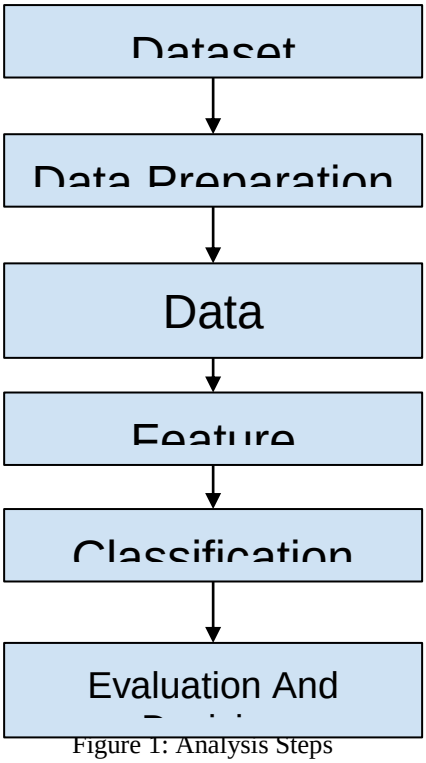

Figure 1: Analysis Steps

3.2 Data Collection Procedure/Dataset Utilized:

We get the dataset ‘Customer churn dataset’ from Kaggle. This dataset has 21 Attributes among those 21 attributes 20 are the selection attribute and 1 is the target attribute. The target attribute is Label(Churn /not-Churn).

Dataset Preparation:

In our dataset we dropped the unnecessary column CustomerId. We take all the data of the data set in numerical form.Then we scale the dataset data. We use MinMaxScale for scaling the dataset.We balance our dataset by using smote.Then we divided our dataset into two parts: one for training and one for testing.The data is split into two parts, with 80% used as training data and the remaining 20% used as testing data.There are about 5625 training data and 1407 testing data.The dataset has 26 training columns and 1 target column. Our target column is the crushed prediction which is true or false.

3.3 Statistical Analysis:

The purpose of customer churn prediction is to figure out which customers are likely to stop using the company's services and take the required steps to reduce customer churn. Statistical analysis is a key component of customer churn prediction. It involves using statistical techniques to analyze historical customer data to identify patterns and relationships between customer behavior and customer churn. Statistical analysis can be used to identify the most important factors that influence customer churn, and help to identify customers at risk of churning. By using statistical analysis, businesses are able to create models that can anticipate customer churn. These models can be used to recognize customers who might be planning to leave the company soon and can help them to develop strategies to keep them with the company. Furthermore, it can be used to figure out which customer segments are more likely to churn, so that the organization can target them with the suitable retention approaches.

3.4 Proposed Methodology/Applied Mechanism:

Applied Models: We define the seven different applied machine learning approaches for our problem:

1)MLPClassifier

2)SVM

3)KNN

4) Decision Tree Classifier

5) RandomForestClassifier

6)GaussianNB

7)Deep neural network

MLP Classifier:

The Multi-Layer Perceptron (MLP) Classifier is a type of artificial neural network that is suitable for use in customer churn prediction. It is capable of performing complex nonlinear computations and is well-suited to problems where the underlying data are highly structured and contain many meaningful features. The MLP Classifier is trained using a supervised learning algorithm and is able to identify patterns in the data that can be used to accurately predict customer churn. By leveraging the MLP Classifier, businesses can develop churn prediction models that are more accurate and reliable than traditional methods of customer attrition analysis.

MLPClassifier (Multilayer Perceptron Classifier) is a supervised learning algorithm that is used for both classification and regression tasks. It is a feedforward artificial neural network model that is trained using backpropagation. The MLPClassifier can be used to predict customer churn by training the model on customer data such as demographics, billing information, customer service interactions, and other related factors. This model can then be used to predict the likelihood of a customer to churn or remain loyal to the company.

The equation used to calculate the output of the MLP classifier is as follows:

Output = Activation Function(Weighted Inputs + Bias)----------(i)

MLPClassifier is a powerful tool for customer churn prediction due to its ability to find complex non-linear relationships between customer attributes and churn probability. It has a high capacity for handling large datasets with thousands of features, and its layers of neurons allow for more accurate modeling of customer behavior. Additionally, MLPClassifier is faster to train and has better results on noisy data than other classification models such as Support Vector Machines.

SVM:

Support Vector Machine (SVM) is a powerful and versatile machine learning algorithm that can be used for customer churn prediction. It is based on the concept of finding a hyperplane to separate data points of different classes. It is a supervised learning algorithm that can be used for both classification and regression tasks. In the case of churn prediction, it can be used to identify customers who are likely to churn in the near future, based on their past behaviors. The algorithm works by taking a set of customer data points and plotting them on a graph based on the features they possess. It then tries to find the optimal hyperplane that best separates the data points into the two groups, customers who are likely to churn and customers who are not. The algorithm then uses this hyperplane to predict the churn rate of new customers. By using SVM for customer churn prediction, companies can accurately identify customers who are likely to churn, and take appropriate actions to retain them.

We load important libraries. Then we import the dataset and separately extract the X and Y variables.Next divide the dataset into train and test.We are creating the SVM classifier model from scratch.Then the SVM classifier model is being fit.Next ,Coming up with predictions.

The SVM algorithm equation describes a decision boundary (or hyperplane) in a multidimensional space that separates data points of different classes.

The equation for a linear Support Vector Machine (SVM) is:

$w \cdot x + b = 0$----------(i)

where w is the weight vector,
x is the feature vector, and
b is the bias.
where w is a vector of weights, x is a vector of feature values, and b is a bias. The decision boundary is determined by the weights and bias, and the distance of the decision boundary to the origin of the coordinate system is determined by the bias.

KNN:

K-Nearest Neighbors (KNN) is a supervised machine learning algorithm used for the purpose of classification and regression. It is a powerful and popular algorithm for predicting customer churn. In customer churn prediction, KNN is used to identify patterns in customer behavior that indicate a likelihood of them leaving the company. KNN works by using the customer's current behavior as a proxy for their future behavior, and then predicting the probability of churn based on the similarity of the customer's behavior to other customers who have previously churned. KNN is an effective tool for customer churn prediction because it can identify similar customers who have left in the past, and then use those customers' behavior to predict the likelihood of churn for the current customer.

The K-Nearest Neighbors (KNN) algorithm is based on the distance measure between the test sample and the training samples:

$$Knn = 1/ (1 + (distance/ k)) \quad \text{----------(i)}$$

Where,

Knn = Weighted average of the neighbors

distance = Euclidean distance between the data points

k = Number of nearest neighbors used to predict the outcome

KNN is an effective algorithm for customer churn prediction, as it uses the customer's historical data to make predictions. It is a non-parametric algorithm that relies on the similarity of customer data points to make predictions. KNN is good for customer churn prediction because it can use data from a customer's past behavior to predict future behavior, which can be helpful in understanding what factors may lead to customer churn. Additionally, KNN is easy to implement and can be used to quickly create customer churn predictions.

Decision Tree Classifier:

A Decision Tree Classifier is a supervised learning algorithm that uses a tree-like structure to classify different objects. It uses a decision tree, which is composed of decision nodes, to classify a given input into one of several output classes. At each decision node, the algorithm determines which attribute to split the data on. The algorithm then assigns the data to one of the classes based on the value of the attribute. The goal of the decision tree classifier is to correctly classify the data into one of the output classes.

Here it is used to identify patterns in customer data, such as customer demographics and usage history. The algorithm works by constructing a tree-like structure, where each node represents a decision based on the data. In each node, the algorithm splits the data into two or more subsets based on the most important attribute. The algorithm then recursively evaluates the subsets to reach the most accurate prediction. The prediction is based on the probability of a customer remaining or leaving the service. The Decision Tree Classifier uses a variety of metrics to evaluate the data, such as Gini Index and Information Gain. The Gini Index measures the purity of the subset while Information Gain measures the relative importance of the attribute in the prediction. The algorithm can also be used to identify the most important factors in customer churn, allowing companies to take action to reduce customer attrition.

Decision Tree Classifier:

Prediction = (Probability of a customer churning given a set of features) * (Probability of the customer not churning given a set of features)----------(i)

= (P(Churn | Feature1, Feature2, ..., FeatureN)) * (1 - P(Churn | Feature1, Feature2, ..., FeatureN))

Random Forest Classifier:

A Random Forest Classifier is a powerful machine learning algorithm used for predicting customer churn. It works by constructing a multitude of decision trees, each one using a different combination of the input variables, and then combining their results to form a final prediction. The trees in a Random Forest Classifier look at different combinations of the input variables and use a voting system to determine the final prediction. This helps to reduce the variance of the model and produce a more accurate prediction. The Random Forest Classifier is particularly effective when the data is complex, as it is able to capture relationships between variables that may be difficult to detect with a single decision tree. Additionally, by combining the results of multiple trees, the Random Forest Classifier is able to make more accurate predictions than a single decision tree. As such, the Random Forest Classifier is a powerful predictive tool for customer churn prediction.

Random Forest Classifier algorithm equation is:

$P(Y|X) = 1/K \Sigma T(x)$ ----------(i)

where,

P(Y|X) = probability of the customer churn (Y) given attributes (X)

K = number of trees

T(x) = prediction by a single tree for a given x

We take samples randomly from a data set or training and testing sets. Then, we build a decision tree with each set of training data. Afterwards, we average the decision trees to find the most

popular prediction result. Finally, the prediction result that received the most votes is chosen as the final forecast result.

GaussianNB

GaussianNB is a machine learning algorithm used for classification problems. It is a supervised learning algorithm, meaning it uses labeled data to learn how to classify data points. The algorithm uses Bayes' Theorem to determine the probability of a data point belonging to a particular class.

GaussianNB can be used for customer churn prediction as it can identify patterns in customer behavior and use them to classify customers as “likely to churn” or “unlikely to churn”. The algorithm can also use customer information such as age, gender, location, product usage, etc. to better understand customer behavior and make more accurate predictions.

Y = GaussianNB(X)----------(i)

Where:

Y = Customer Churn Prediction
X = Features (e.g. customer demographics, usage data, etc.)

We are first loading the initial libraries, then importing and exploring the dataset, followed by visualizing it. Assuming that all the continuous variables associated with each feature have a uniform distribution, we can use a method called Gaussian Naive Bayes. This is also known as the normal distribution and if our data is continuous, we can use this formula to calculate the probability of likelihoods.

Deep neural network

A deep neural network is a powerful machine learning technique used to detect patterns in customer data and make predictions about customer churn. It can be used to identify factors that may lead to customer churn and build predictive models based on those factors. The neural network can then be used to make predictions about customer churn using the input data. The predictive models can be used to classify customers into different churn categories, such as high, moderate, and low churn. Additionally, the neural network can be used to make predictions about the future behavior of customers and provide insights into customer behavior.

Deep Neural Networks (DNNs) are a type of artificial neural network (ANN) that has multiple hidden layers between the input and output layers. Each hidden layer uses nonlinear activation functions to transform the input data into a more complex representation. This allows the network to learn complex relationships between inputs and outputs, making them ideal for tasks such as customer churn prediction. DNNs can also capture both linear and nonlinear patterns in data, allowing them to model data more accurately than traditional methods such as logistic regression or decision trees. DNNs are also able to learn from large amounts of data, making them well suited for tasks with large datasets. DNNs can be trained using backpropagation, which involves adjusting the weights of connections between the various layers in order to minimize errors in the output. The result is a model that is capable of making accurate predictions about customer churn.

Deep neural networks are used for customer churn prediction as they are able to capture complex non-linear relationships between customer data and the likelihood of customer churn. Deep neural networks can also capture temporal dependencies between customer data points, allowing for the detection of subtle patterns which are difficult to detect using traditional machine learning algorithms. Deep neural networks have been shown to outperform traditional machine learning algorithms in customer churn prediction tasks, making them a powerful tool for predicting customer churn.

3.5 Implementation Requirements:

1. Data Collection: Collecting customer data such as customer demographics, customer behavior, customer transactions, etc.
2. Data Cleaning & Pre-processing: Cleaning, formatting, and normalizing the collected data to make it ready for the predictive model.
3. Feature Engineering: Identifying significant characteristics from the data that can be employed to anticipate customer churn.
4. Model Selection: When it comes to predicting customer churn, choosing the right machine learning or deep learning model is essential. It is important to consider the type of data available, the complexity of the problem, and the desired accuracy of the model in order to select the model that will best suit the needs of the project. Additionally, the amount of time, resources, and expertise available for the project should be taken into consideration when selecting the appropriate model.
5. Model Training and Evaluation: Training the selected model on the pre-processed and engineered data and evaluating its performance.
6. Model Deployment: Deploying the trained model in a production environment and integrating it with existing systems.

Chapter 4
Experimental Results and Discussion

4.1 Experimental Setup:

Gather all relevant customer data such as customer demographics, past behavior, and current activity.Clean the data, perform any necessary transformations, and encode categorical features. Creating various machine learning models, such as decision trees, logistic regression, and random forests, can be beneficial in many different ways.It is important to measure the performance of each model using metrics such as accuracy, precision, recall, and F1 score. After assessing each model, the best one should be chosen based on the evaluation metrics.Deploy the model in a production environment and monitor its performance over time.

4.2 Experimental Results & Analysis:

Table 1 shows that the RandomForestClassifier is the best performer among these machine learning and Deep learning algorithms, with an accuracy of 84.00%. With a testing accuracy of 83.74%,Random Forest Classifier was found to be the best performer. So we used RandomForestClassifier to create our prediction algorithm and improved it to improve prediction accuracy. Figure (2) depicts the performance analysis of benchmark models.

Table1: Accuracy table of Machine Learning Algorithms

Model	ACCURACY	Training Accuracy	Testing Accuracy
MLP Classifier	82.00%	87.40%	82.38%
SVM	81.00%	83.08%	80.64%
KNN	80.00%	85.86%	79.96%
Decision Tree Classifier	76.00%	99.87%	76.38%
Random Forest Classifier	84.00%	99.87%	83.74%
Gaussian NB	77.00%	76.48%	77.09%
Deep neural network	81.00%	85.41%	81.36%

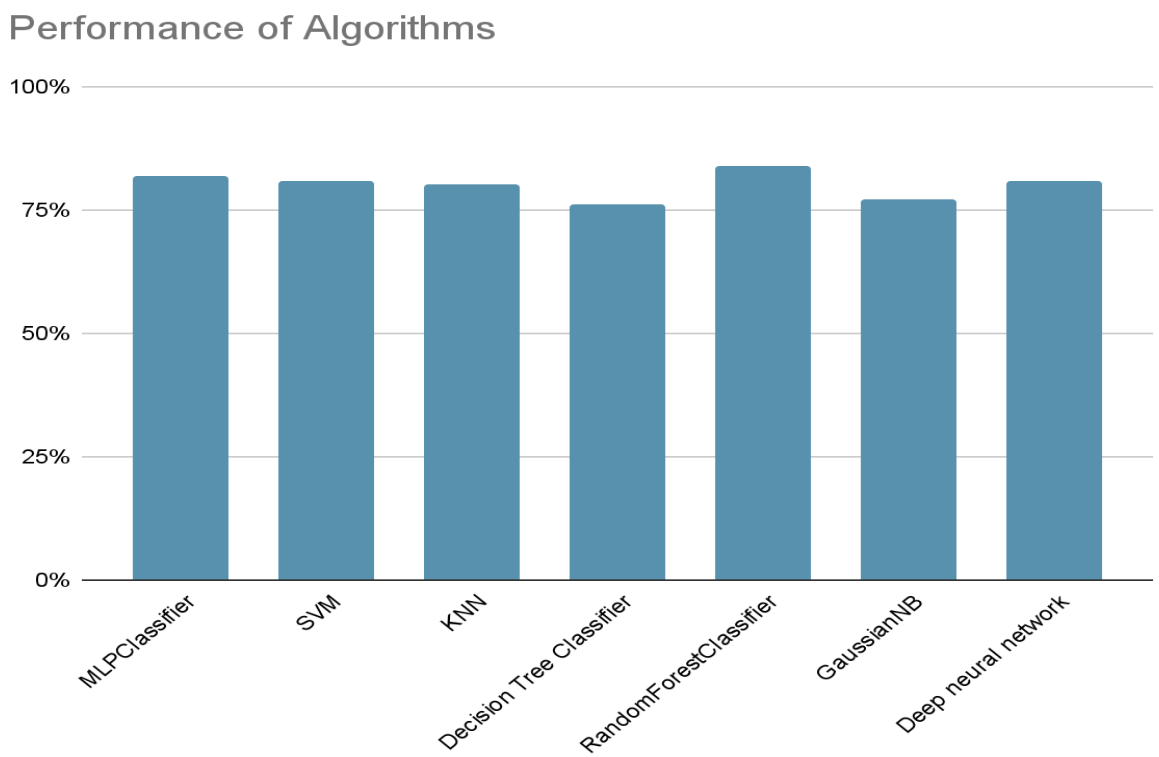

Figure 2: Performance analysis of machine learning models

We obtained precision values (figure:3) (Precision is the model's ability to predict a specific category.) for Churn respectively MLP Classifier (80.00%), SVM (78.00%), KNN (75.00%), Decision Tree Classifier (76.00%), Random Forest Classifier (81.00%), GaussianNB (74.00%), Deep neural network (79.00%) and not churn MLP Classifier (85.00%), SVM (83.00%), KNN (87%), Decision Tree Classifier (77.00%), Random Forest Classifier (86.00%), Gaussian NB (80.00%), Deep neural network (85.00%). It would be fantastic if the machine made the correct decision.

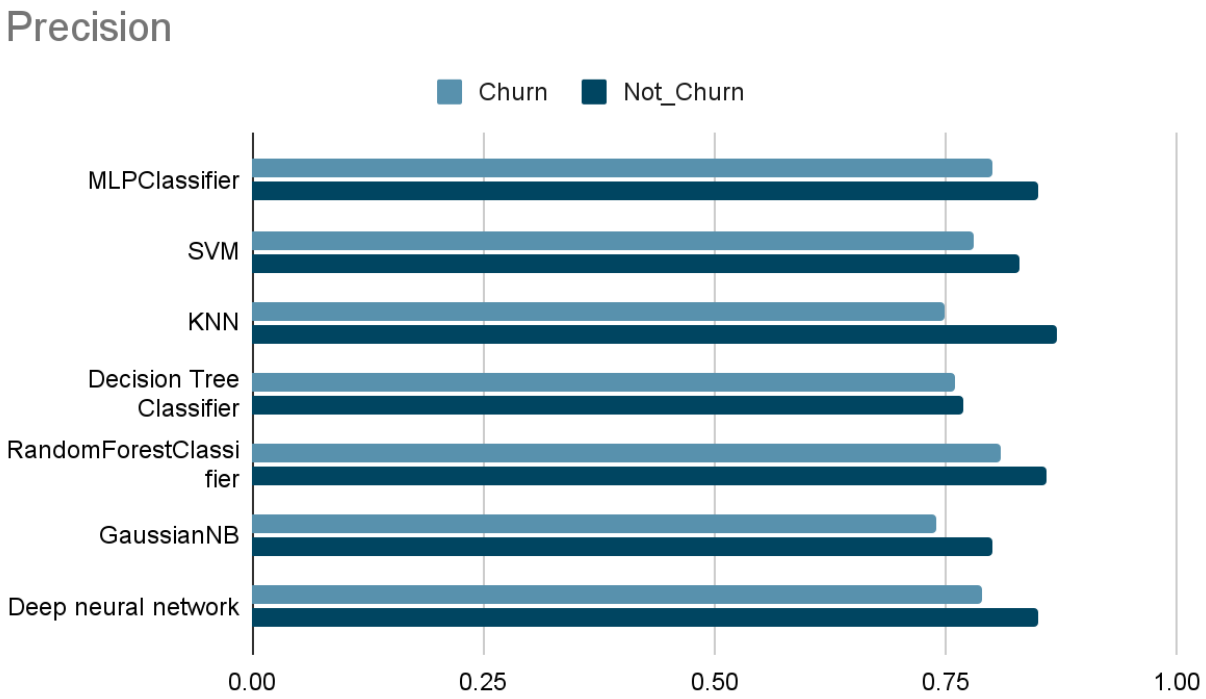

Figure 3: Precision values from different models

We obtained recall values (figure :4) (it is the measure of our model correctly identifying True Positives) for Churn respectively MLPClassifier(86.00%), SVM (84.00%), KNN (89.00%), Decision Tree Classifier(76.00%), RandomForestClassifier(87.00%), GaussianNB (82.00%), Deep neural network(86.00%) and not_churn respectively MLPClassifier(79.00%), SVM (77.00%), KNN (71.00%), Decision Tree Classifier(76.00%), RandomForestClassifier(80.00%), GaussianNB (72.00%), Deep neural network(77.00%). This will be extremely helpful in making the best machine decision.

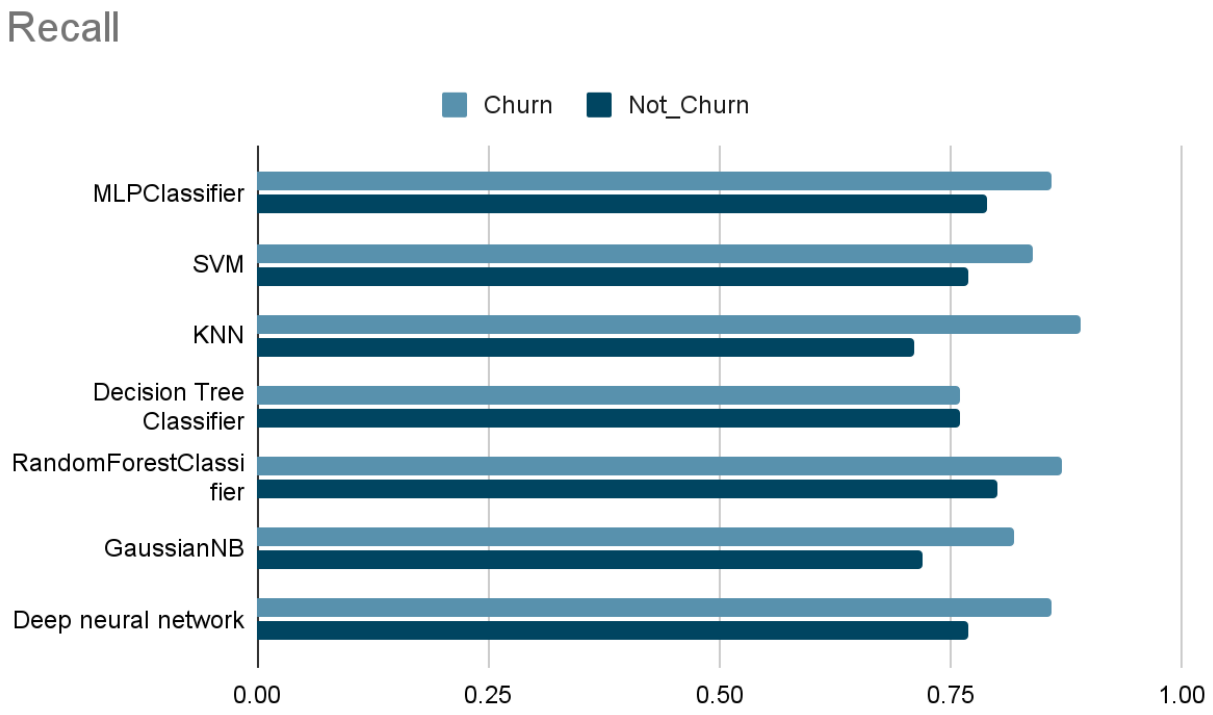

Figure 4: Recall values from different models

4.3 Roc_Curve:

The ROC curve(fig:5) depicts the trade-off between sensitivity and specificity: The percentage of positive cases that were accurately identified is how sensitive a system is. The proportion of unfavorable cases that were correctly categorized serves as a gauge of specificity.

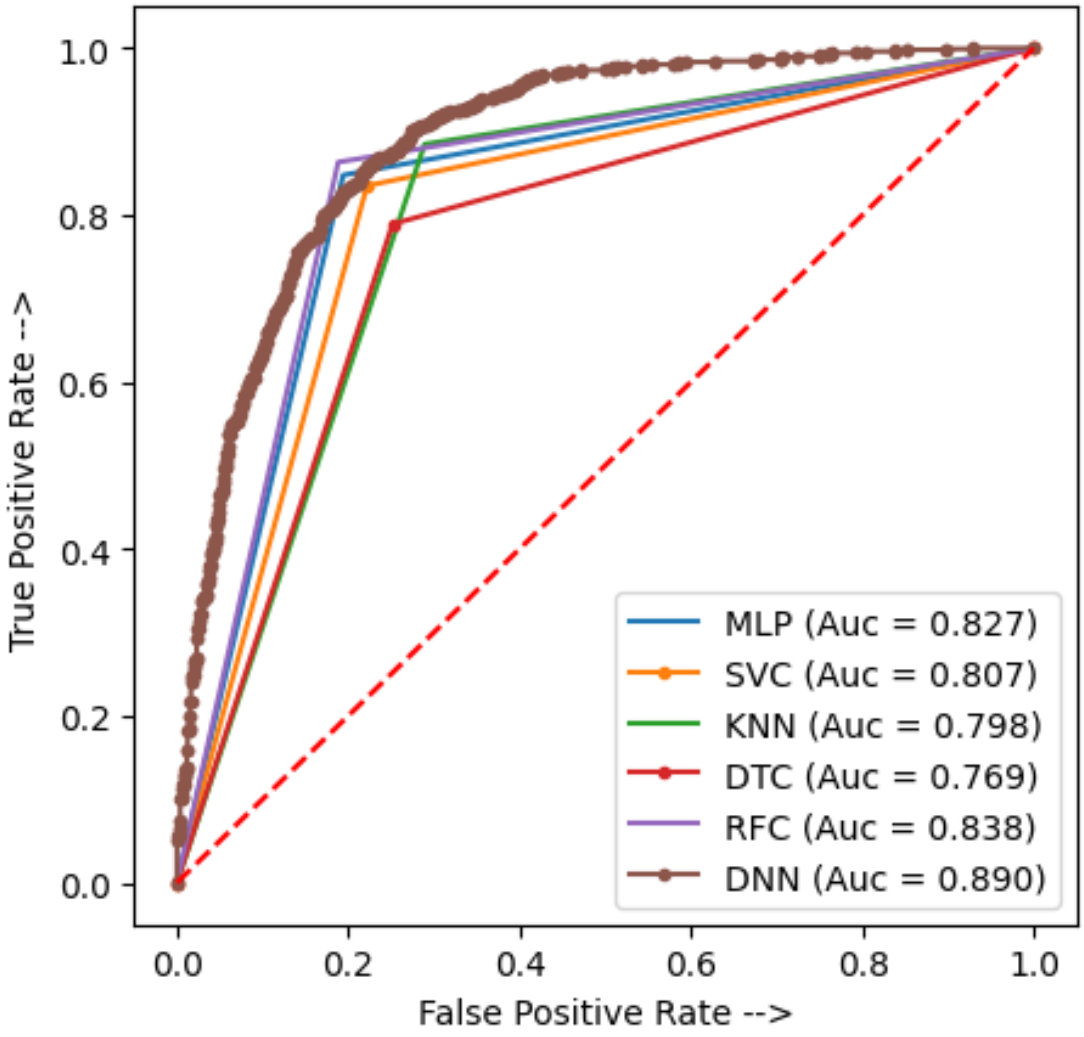

Figure 5: Roc_curve of different models

4.4 Discussion:

The customer churn prediction result is a valuable tool for companies to identify customers that are at risk of churning. By identifying churners, businesses can take preemptive action to retain those customers and reduce their churn rate. The result of customer churn prediction can also help businesses understand what factors are driving customer churn, allowing them to adjust their strategies to reduce churn. Furthermore, businesses can use customer churn prediction to target customers for more personalized offers and promotions, as well as to create loyalty programs to reward customers for their loyalty. Ultimately, a customer churn prediction result can help businesses to improve customer retention and increase customer lifetime value.The RandomForestClassifier is the best performer among these machine learning and Deep learning algorithms, with an accuracy of 84%. With a testing accuracy of 83.74%,RandomForestClassifier was found to be the best performer. So we used RandomForestClassifier to create our prediction algorithm and improved it to improve prediction accuracy.

Chapter 5
Summary & Conclusion

5.1 Summary of the Study:

The study of customer churn prediction is a research topic that focuses on predicting when a customer is likely to end a business relationship. This is important for businesses because it allows them to identify and target customers who are likely to leave before it becomes too late.

The study of customer churn prediction involves a number of different techniques and methods. These include predictive analytics methods such as logistic regression and decision trees, as well as machine learning methods such as support vector machines and neural networks. In addition, customer segmentation and customer relationship management (CRM) techniques can be used to better understand the customer behavior that is associated with churn and to create targeted interventions that can help reduce churn.

The study also investigates the important factors that influence customer churn, such as customer demographics, product features, and customer service. In addition, research has been conducted to understand how customer satisfaction impacts churn and how customer loyalty can be built and maintained. Finally, the study also investigates the effectiveness of various strategies for reducing churn.

Overall, the study of customer churn prediction is an important research topic that can help businesses better understand and predict customer behavior, which can ultimately lead to improved customer retention.

5.2 Conclusions:

The conclusions of customer churn prediction are that customer churn can be accurately predicted by using a combination of machine learning models and data-driven analysis. Specifically, customer churn can be predicted by leveraging customer demographic and transactional data, as well as customer behavior data. Additionally, customer segmentation can help identify customers who are more likely to churn, and some predictive models can be used to

identify the risk of churn for individual customers. By leveraging these techniques, businesses can take proactive steps to reduce customer churn.

5.3 Implication for Further Study

In order to further study customer churn prediction, it would be beneficial to incorporate more data sources into the analysis. Such additional sources could include customer surveys, customer feedback, and customer segmentation. Additionally, machine learning methods such as decision trees, random forests, and support vector machines could be used to develop more accurate models for predicting customer churn. Finally, the use of more sophisticated metrics and techniques for evaluating the performance of customer churn models would be beneficial.

5.4 Contributions

This thesis will include background information and a study of the relevant literature on the impact of customer analytics on customer attrition. This will be done by splitting the subject into two sections, one focused on customer analytics and the other on customer turnover, along with their respective subsections. Additionally, a full examination of the most current advancements in data mining and analysis, particularly those that pertain to the utilization of customer data, which we characterize as customer analytics, will be provided.

References:

[1] Saran Kumar, A., & Chandrakala, D. (2016). A survey on customer churn prediction using machine learning techniques. *International Journal of Computer Applications*, *975*, 8887.

[2] Xia, G. E., & Jin, W. D. (2008). Model of customer churn prediction on support vector machine. *Systems Engineering-Theory & Practice*, *28*(1), 71-77.

[3] Rahman, M., & Kumar, V. (2020, November). Machine learning based customer churn prediction in banking. In *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)* (pp. 1196-1201). IEEE.

[4] de Lima Lemos, R. A., Silva, T. C., & Tabak, B. M. (2022). Propension to customer churn in a financial institution: A machine learning approach. *Neural Computing and Applications*, 1-18.

[5] Matuszelański, K., & Kopczewska, K. (2022). Customer Churn in Retail E-Commerce Business: Spatial and Machine Learning Approach. *Journal of Theoretical and Applied Electronic Commerce Research*, *17*(1), 165-198.

[6] Wael Fujo, S., Subramanian, S., & Ahmad Khder, M. (2022). Customer Churn Prediction in Telecommunication Industry Using Deep Learning. *Information Sciences Letters*, *11*(1), 24.

[7]Kaur, I., & Kaur, J. (2020, November). Customer churn analysis and prediction in banking industry using machine learning. In *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)* (pp. 434-437). IEEE.

[8]Kavitha, V., Kumar, G. H., Kumar, S. M., & Harish, M. (2020). Churn prediction of customer in telecom industry using machine learning algorithms. *International Journal of Engineering Research and Technology (IJERT)*, *9*(5), 181-184.

[9] He, Y., Xiong, Y., & Tsai, Y. (2020, April). Machine learning based approaches to predict customer churn for an insurance company. In *2020 Systems and Information Engineering Design Symposium (SIEDS)* (pp. 1-6). IEEE.

[10] Gaur, A., & Dubey, R. (2018, December). Predicting customer Churn prediction in telecom sector using various machine learning techniques. In *2018 International Conference on Advanced Computation and Telecommunication (ICACAT)* (pp. 1-5). Ieee.

[11] S. Kim, and H. Lee. "Customer Churn Prediction in Influencer Commerce: An Application of Decision Trees." Procedia Computer Science, vol.199, pp. 1332-1339, 2022.

[12] K. Geetha, P. Tomar, and A. Jain. "Customer Churn Prediction." *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* vol. 13, no. 2, pp. 944-948, 2022.

[13] F. O.Koçoğlu, and T. Özcan. "A grid search optimized extreme learning machine approach for customer churn prediction." *Journal of Engineering Research* (2022).

[14] Z.Gao, Y. Shu, X. Zhang, and X. Zhang. "Telecommunication customer churn prediction based on multiple machine learning algorithms." In *5th International Conference on Computer Information Science and Application Technology (CISAT 2022)*, vol. 12451, pp. 1135-1139. SPIE, 2022.

Appendices:

https://colab.research.google.com/drive/1ZivLCjn7snL2rvcP0FrNd7cCCU7_FS3-?usp=sharing
https://drive.google.com/file/d/115hMTK8OFXQNuI6hOh2Kmr7Q13NJhlUN/view?usp=sharing
https://drive.google.com/drive/folders/1x-fNPspInDDrsWdbJN3aWsK3KB-4Sueu?usp=sharing
https://github.com/Prattoypaul/deeplearning/blob/735acb77abcb7c45b853181fc8cbc986ab283ded/Churn.ipynb

Customer Churn Prediction with Machine Learning Approaches

ORIGINALITY REPORT

27%	24%	11%	18%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	8%
2	Submitted to Daffodil International University Student Paper	4%
3	Submitted to Sheffield Hallam University Student Paper	1%
4	www.mdpi.com Internet Source	1%
5	link.springer.com Internet Source	1%
6	www.researchgate.net Internet Source	1%
7	ejournal.bsi.ac.id Internet Source	1%
8	www.ijisae.org Internet Source	1%
9	"International Conference on Innovative Computing and Communications", Springer	1%