COMPARATIVE ANALYSIS OF MACHINE LEARNING TECHNIQUES FOR PREDICTING CUSTOMER CHURN IN THE TELECOM INDUSTRY

BY

Sazzad Hossain

ID: 0242310005103005

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

Supervised By

Dr. S. M. Aminul Haque

Professor & Associate Head

Department of Computer Science and Engineering

Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JULY 2024

APPROVAL

This thesis titled "**Comparative Analysis of Machine Learning Techniques for Predicting Customer Churn in the Telecom Industry**", submitted by Sazzad Hossain, ID No: 0242310005103005 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 M.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 06-07-2024.

BOARD OF EXAMINERS

Narayan Ranjan Chakraborty
Associate Professor and Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman

Md. Sadekur Rahman
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dr. Ohidujjaman, PhD
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Mr. Sadat Hossain
Data Scientist
Risk Management Division,
BRAC Bank Limited

External Examiner

DECLARATION

I hereby declare that, this thesis has been done by me under the supervision of **Dr. S. M. Aminul Haque, Professor & Associate Head**, Department of CSE, Daffodil International University. I also declare that neither this thesis nor any part of this thesis has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Dr. S. M. Aminul Haque
Professor & Associate Head
Department of CSE
Daffodil International University

Submitted by:

Sazzad Hossain

Sazzad Hossain
ID: 0242310005103005
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, I express my heartiest thanks and gratefulness to almighty God for His divine blessing makes it possible to complete the final thesis successfully.

I am really grateful and wish my profound indebtedness to **Dr. S. M. Aminul Haque, Professor & Associate Head, Department of CSE**, for his deep knowledge & keen interest in the field of Machine Learning to carry out this thesis. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this thesis.

I would like to express my heartiest gratitude to **Dr. Sheak Rashed Haider Noori, Professor and Head, Department of CSE**, for his kind help to complete my thesis and also to other faculty members and the staff of CSE department of Daffodil International University.

I would also like to thank my entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, I must acknowledge with due respect the constant support and passion of our parents.

ABSTRACT

Customer churn prediction is a critical aspect of the telecommunications industry, significantly impacting a company's profitability and customer retention strategies. This thesis aims to develop an effective predictive framework using various machine learning algorithms. The study utilizes the IBM Telco dataset, which contains a class imbalance of 26.54% churned customers and 73.46% non-churned customers, across 33 features. The research focuses on evaluating the performance of seven machine learning techniques: Gradient Boosting Classifier, K-Nearest Neighbors (KNN), Logistic Regression, Naive Bayes, Random Forest, Support Vector Machine (SVM), and XGBoost. Grid Search Cross Validation, with K-Fold sizes of 5 and 10, is employed to optimize hyperparameters and ensure robust model selection. Data preprocessing techniques such as handling class imbalance, feature selection, and engineering are applied to enhance model accuracy. Experimental results demonstrate that Random Forest and XGBoost achieved the highest accuracy of 84% with a 5-fold cross-validation, while Naive Bayes exhibited the lowest performance at 77%. The Random Forest algorithm also recorded the best accuracy of 86% with a 10-fold cross-validation. The thesis addresses several challenges, including class imbalance, overfitting. This research contributes to the field by offering a comparative analysis of machine learning techniques and their effectiveness in predicting customer churn, ultimately aiding telecom companies in devising better customer retention strategies. This study's findings will inform industry practitioners and researchers about the most effective machine learning practices for churn prediction, emphasizing the critical role of accurate and reliable models in maintaining customer loyalty and reducing churn rates.

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 Introduction

Customer churn is a pivotal metric for businesses, representing the rate at which customers cease their engagement with a company's products or services within a defined timeframe. It serves as a critical indicator of customer satisfaction and loyalty, reflecting how well a company meets customer needs and expectations. High churn rates suggest underlying issues such as product dissatisfaction, poor customer service, or ineffective marketing strategies, which can adversely affect revenue and profitability [1]. While a low churn ratio means people are not leaving the service and the company is growing, high churn ratio is bad for any business because only acquiring new customers can be very expensive, and as soon as the company loses focus on how to retain customers, they die out and the company turns its back to those who choose their business. Unless churn is under control, retentive measures known as retention strategies will be put in place, with a singular goal of improving customer experience and mitigate churn and ensure a constant long-lasting healthy business relationship. By surveying the reasons for churn, we can have a great view of customers' needs and wants, for data-driven decisions that keep us competitive in a fast-growing environment. Customer churn usually expressed as the number of customers lost during a certain period divided by the number of customers total from the start of the period. Churn analysis is the key for businesses that wish to have a clear understanding of customer behavior and improve products and by leveraging churn analysis insights, businesses can proactively address customer needs, strengthen relationships, and drive sustainable growth and resilience in competitive markets [2].

The valuable insights derived from churn analysis help businesses understand why customers are leaving or staying. The pattern and trend analysis of consumer behavior helps in making unbiased appraisal on where the firm may be underperforming or overperforming. This awareness of weaknesses and strengths empowers organizations to make the right decisions about product improvements, feature enhancements, and service adjustments for better customer satisfaction. Not only does this show why customers leave but also opens up opportunities for enhanced interaction with existing customers. Thus, understanding the influencing factors of churn enables companies to fit their

communication strategies towards customer concerns, timely support provision, relevant content delivery or offers resonating to their target groups. Improved communication can strengthen customer relationships, increase loyalty, and ultimately reduce churn [2].

By analyzing historical churn data and identifying key indicators or predictors of churn, businesses can develop predictive models to forecast future churn rates. These predictive insights enable proactive intervention strategies to prevent churn before it happens. Whether through targeted retention campaigns, personalized offers, or proactive customer support, businesses can leverage churn analysis to anticipate customer needs, address potential issues, and enhance overall satisfaction, thereby reducing future churn and maximizing customer lifetime value [2]. During times of crisis or turbulence, such as economic downturns, market disruptions, or unexpected events, churn analysis can serve as a valuable strategic tool. By closely monitoring churn rates and understanding customer behavior dynamics, businesses can quickly adapt their strategies, prioritize resources, and implement targeted initiatives to retain customers, mitigate churn, and safeguard revenue streams. Churn analysis provides actionable insights to navigate challenges effectively, maintain customer trust, and emerge stronger from crises.

Churn prediction is of great importance for profit-oriented business organizations. It entails predicting who are those customers who are most likely to cancel their subscription or stop using the provided services. Predicting churn is an important factor, because it is more expensive to gain a new client than to retain an existing client. Once a client is identified to be at-risk, additional marketing actions can be done for each individual customer to improve the likelihood of the client to stay. Customer churn is an issue for organizations across various business sectors and it signifies an investment lost due to the time and resources needed to acquire a new client. Being able to predict which client will leave next and providing an additional incentive for them to stay can lead to saving on the part of the business. Knowing which factors affect the clients staying or leaving can be useful for devising strategies to retain customer loyalty and operational strategies related to retention. Churn prediction is indeed vital for business organization that offers subscription-based services. Even the slightest change in the churn rate can have an intense multiplicative effect on the balance sheet. The primary objective is to determine whether a customer is likely to leave within a specific timeframe, thereby framing churn prediction as a binary classification task with outcomes of "Yes" or "No" [3]. This predictive capability enables

businesses to proactively address customer attrition, optimize retention efforts, and safeguard long-term profitability.

1.2 Motivation

Customer churn is the number or percentage of a company's customers who stop using the company's products or services during a particular period of time. This is a business issue that could indeed affect a company financially. It is a measure in the assessment of customer satisfaction and loyalty. The reasons for churn, once known, can help a business take proactive measures to prevent churn and avoid revenue loss. Churn management can help companies maintain their customers and, through proper management and focused efforts to contain retention efforts, can convert dissatisfied customers into brand advocates. Customer churn is more than losing a customer. It is a substantial financial blow for businesses that could cause a ripple effect on revenue, market share, and brand reputation. It is a direct translation into declining sales and revenue. Departure of customers means the loss of the recurring revenue from them and potential revenue from newly added customers. This could create a dent in the company's financial performance both immediately and over the long term.

If a customer leaves, the company loses its share of that customer's spend to the competition. If it consistently has higher churn, the competition will be its opportunity loss. But more churn also means that the competitors gain added motivation to pick up the clients sandwiched between them – clients who are already angry and ready to tell that anger to all their friends in the form of negative word of mouth. More and more threads on 'worst customer service' will start being connected to the business. And more and more market share will take a steep dive. So, churn involves lower revenue, sure. But it also involves revenue loss – both because new customers may not be easy to find or cheap to acquire – sometimes a lot more costly – and because the company's brand may start picking up a reputation of poor customer experience.

With these enormous financial ramifications, companies need to create tactics that work for both customer retention and churn reduction. Churn can be reduced by funding proactive client retention strategies including customer support, loyalty programs, and personalized customer experiences. Businesses may preserve their income streams, market share, and brand reputation by putting customer retention efforts first and putting churn reduction tactics into practice. This will ultimately lead to long-term financial stability and growth.

1.3 Research Objective

The primary objective of this research is to develop and evaluate a comprehensive framework for predicting customer churn in the telecommunications industry using various machine learning algorithms. By employing techniques such as K-Fold Cross-Validation for hyperparameter optimization, this study aims to identify the most accurate and reliable models for churn prediction. The goal is to enhance the understanding of customer churn dynamics and provide actionable insights for improving customer retention strategies. The specific objectives are to:

- Address the class imbalance in the dataset to ensure unbiased model performance.
- Explore the impact of feature selection and engineering techniques on model performance, with a focus on identifying the most relevant predictors of churn.
- Optimize hyperparameters for each machine learning algorithm using Grid Search Cross-Validation.
- Prevent model overfitting and ensure models generalize well to unseen data.
- Evaluate the performance of following machine learning techniques: Gradient Boosting Classifier, K-Nearest Neighbors, Logistic Regression, Naive Bayes, Random Forest, Support Vector Machine, and XGBoost.

By addressing these objectives, this research aims to contribute to the advancement of churn prediction methodologies and provide practical guidance for businesses seeking to mitigate customer churn and enhance customer lifetime value in competitive market environments.

1.4 Research Question

To achieve the research objectives, the study will address the following research questions:

- How can class imbalance in the telecom customer churn dataset be effectively managed to improve model performance?
- What are the most relevant features for predicting customer churn, and how can they be engineered and selected effectively?
- What are the most effective data preprocessing methods for improving the quality of the telecom customer churn dataset and enhancing model performance?
- Which machine learning algorithm provides the highest accuracy and reliability in predicting customer churn when optimized with K-Fold Cross-Validation?
- How can hyperparameter tuning be optimized to enhance the performance of each machine learning algorithm?

1.5 Report Layout

The report has total 5 Chapters which will be followed given by instructions:

In Chapter 1, introducing the thesis, exploring the significance of customer churn analysis and prediction in detail. I thoroughly examined the research question, the motivations driving this study, the rationale behind my approach, and the anticipated outcomes.

In Chapter 2, I conducted a comprehensive review of existing literature concerning customer churn prediction. This section encompassed an analysis of various authors' methodologies, limitations encountered, results obtained, and methods employed. Additionally, I explored the research scope and addressed the challenges encountered in this field.

In Chapter 3, I primarily delve into the research methodology. This chapter sheds light on the data collection and pre-processing procedures, offering insights into data statistics, algorithms employed, and the proposed workflow.

In Chapter 4, I primarily present the results of my research. Here, I showcase the outcomes of all the algorithms utilized in this study, including F1 Score, Accuracy, Precision, Recall.

In Chapter 5, the focus shifts to the discussion and conclusion of my research. Here, I engage in a thorough examination of our findings and draw conclusions from them. I also explore potential future directions for research and acknowledge the limitations inherent in my work.

CHAPTER 2
BACKGROUND

2.1 Introduction

In today's competitive landscape, businesses face the ongoing challenge of customer churn, where customers discontinue their relationship with a company. Predicting customer churn is paramount for businesses, especially those operating under subscription-based models like SaaS companies, as churn can significantly impact profitability [8]. Recognizing the importance of customer retention, this study aims to develop a robust churn prediction model using machine learning techniques.

At its core, predicting customer churn requires a deep understanding of customer behavior and preferences. The cost associated with acquiring new customers often outweighs the cost of retaining existing ones, underscoring the financial significance of reducing churn rates [8]. In the realm of subscription-based businesses, such as SaaS companies, mitigating churn becomes even more critical due to the recurring revenue model [9].

This research focuses on analyzing several machine learning models enhanced with K-fold Cross Validation in order to address the problem of churn prediction. The study attempts to improve prediction accuracy and streamline model selection by utilizing cross-validation procedures and sophisticated modeling approaches.

2.2 Related Works

Numerous recent studies have made significant contributions to the field of customer churn prediction. In this section, we examine several methods explored by various researchers for predicting customer churn.

The study titled "Customer Churning Analysis using Machine Learning Algorithms" published in 2023 by KeAi Communications Co, contributes to the ongoing research on customer churn prediction within the domain of machine learning [10]. The previous research by many researchers in customer churn prediction using machine learning has highlighted the performance of machine learning and artificial intelligence in the sector but has also brought the attention that integrated solutions are preferred over individual performances [10]. Particularly, deep learning algorithms have been suggested for image-

based reviews and sentiment analysis, and there has also been a significant campaign to optimize models using feature engineering.

The methodology is designed to work on previous information to churn the customers. Using a 7044 alternative, the 21-attribute dataset taken from Kaggle, the following methodology is implemented. First, the data set is subjected to initial steps of preprocessing and analysis. It is then separated into training and testing sets. Several predictive models in the name of Logistic Regression, K-Nearest Neighbors, Random Gradient Booster, and Random Forest have been utilized. Ensemble techniques have been applied to understand the influence of all these models on the ultimate prediction accuracy, with the preliminary objective of the study being to develop a strong data platform for churn prediction.

The study measures four algorithms using Receiver Operating Characteristics and Area Under Curve, and, according to it, the Stochastic Gradient Booster model has the best output, AUC being 0.84, but the K-Nearest Neighbor model is still slightly weaker, with AUC being 0.781. The study has also noted several shortcomings, among them are substandard data preprocessing, and low hyperparameter standardization. This can hamper the quality and performance of the analysis to great extents. The Grid Search CV method is very time-consuming, and its alternative, the Randomized Search CV, is not able to assure the best hyperparameters in the end; it points out the areas the study to improve.

The paper "Customer Churn Prediction in Telecommunication Industry Using Deep Learning" published in 2022 by Natural Sciences Publishing makes a contribution to the field of customer churn prediction by presenting a deep learning-based solution. It seeks to demonstrate the superiority of their Deep-BP-ANN model in predicting customer churn over traditional machine learning techniques and existing deep learning models, specifically those that make use of holdout or 10-fold cross-validation. Prior studies within this domain have either focused on the use of deep learning techniques to improve customer churn prediction, such as automating feature engineering with Deep Neural Networks (DNNs), or explored applying transfer learning to enhance churn prediction accuracy, but the effect of deep learning in feature creation and its comparison with traditional methods has also been explored as part of the goal to improve churn prediction methodologies [11].

The proposed methodology is based on supervised classification deep learning for telecom customer churn prediction, as opposed to traditional machine learning techniques such as Logistic Regression, Naïve Bayes, K-Nearest Neighbors (KNN), and XGBoost. The process includes data collection, pre-processing, feature engineering, random oversampling to rectify data imbalance, modeling, and evaluation using holdout and K-fold cross-

validation techniques. Feature selection techniques such as Lasso Regularization and feature correlation removal are used to enhance model performance.

A salient point is that this study uses two datasets, namely Cell2Cell and IBM Telco, from Kaggle, with attributes showing customer churn. The Deep-BP-ANN model with feature selection, early stopping, and regularization techniques has outperformed most of the machine learning and available deep learning techniques, thus proving its efficiency in churn prediction on the two datasets.

The paper is on a critical issue in the telecom industry and applies state-of-the-art deep learning techniques to model complex relationships. It recognizes several limitations of deep learning models, including their being computationally intensive, potential model validation issues, and the fact that no exploration was done on other relevant algorithms like random forest and decision trees. The other thing is that the paper lacks a discussion on the hyperparameters selected for the Deep-BP-ANN model, which will make it difficult to either replicate or understand the setup of the model.

2.3 Scope of the problem

Customer churn, defined as the phenomenon in which customers drop their subscription or service, is a significant problem in the telecommunications industry. Detecting and predicting customer churn can help businesses retain a customer base, reduce revenue loss, and maintain market competitiveness. The firms that can predict churn more effectively can devise better strategies for retention and improve customer satisfaction. In this paper, we address the challenging problem of customer churn prediction by applying and assessing seven machine learning techniques: Gradient Boosting Classifier, K-Nearest Neighbors (KNN), Logistic Regression, Naive Bayes, Random Forest, Support Vector Machine (SVM), and XGBoost. We conducted a proper analysis through Grid Search Cross-Validation to highlight the best hyperparameters and also to focus on the best model performance in terms of accuracy, precision, F1 score, and recall. This study will deliver a robust framework for customer churn prediction and benefit the telecommunication firm with sound decision-making in the area of retention.

2.4 Challenges

There are a few underlying challenges of predicting customer churn in the telecommunications industry that need to be addressed for accurate and reliable outcomes: Class imbalance, where the dataset has a lot more non-churned than churned customers,

making the model biased toward the majority class. To mitigate this, techniques like oversampling, under-sampling, or specialized algorithms need to be used. Feature selection and engineering are very important elements that enhance model performance and are identified by a time-consuming and domain expertise-requiring task of finding and creating relevant features. Hyperparameter tuning to machine learning algorithms is essential for the optimization of algorithms, but it is computationally expensive and time-consuming. This becomes a more complex task with methods like Grid Search Cross-Validation. Overfitting is another major challenge. The model should not overfit and should work well on unseen data. For the feature-engineered model to work well, careful validation techniques and regularization are needed. The interpretability of models, specifically more complex ones like Gradient Boosting and Random Forest, will be a key factor in understanding the factors driving customer churn and making the results of the analysis actionable for the business. Preprocessing the data is very important, as there might be missing values, inconsistencies, or possible errors in the dataset. Access to a high-performance computing infrastructure is necessary since the extensive computational resources needed for running several algorithms with extensive hyperparameter tuning will be extensive. Selection of appropriate evaluation metrics, such as accuracy, precision, recall, F1 score, and AUC, to comprehensively evaluate each of the models. By addressing these challenges, this study develops robust and accurate models for predicting customer churn and provides valuable insight and tools for effective customer retention strategies in the telecommunications industry.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introduction

This section serves as an outline of our thesis and the methods employed therein. It provides a comprehensive overview of the data collection efforts and the actionable tactics derived from them. For our purposes, this translates to pinpointed application and actualization of our research objectives. Moreover, it underscores the credibility of our data collection efforts and demonstrates how we intend to leverage them to predict customer churn. Within this section, we will delve into the research methods and tools utilized, offering insights into our approach and the strategies employed to achieve our research goals.

3.2 Research Subject and Instrumentation

Predicting customer churn in the telecommunication sector is significant due to the great impact it may cause in the long run. First, this will affect the revenue streams directly, as it may result in the loss of recurring subscription fees and upselling. Second, the cost of gaining a new customer is usually much higher than the cost of maintaining an existing one. Third, effective churn prediction allows the company to improve the overall experience for the customer by working on the causative factors that may result in dissatisfaction, and this may lead to greater loyalty and retention rates. In a highly competitive market, the capability of predicting and acting on churn gives companies a strategic advantage, enabling them to fine-tune their retention programs and manage resources more efficiently. Advanced machine learning models, such as those used in this study, which are based on the IBM Telco dataset, promise to ensure proactive churn management and, hence, increased business performance. In this paper, a wide set of machine learning algorithms will be used to predict customer churn for the telecommunications industry using the IBM Telco data set. Several well-known algorithms will be implemented in the instrumentation procedure, including the gradient boosting classifier, k-nearest neighbors, logistic regression, naive Bayes, random forest, support vector classifier, and extreme gradient boosting. Each of the algorithms is going to be fine-tuned through meticulous configuration to ensure that it attains maximum predictive power. The reason for choosing these particular machine learning algorithms is that they have a

proven track record of effectiveness in classification tasks, and secondly, they are ideally suited to handling the complexities of churn prediction in large-scale telecom databases.

3.3 Data Description

The dataset used in this research holds 33 columns, representing different features related to customer churn in the telecommunications domain. Some of these attributes include 'CustomerID', which is the identification number for each customer; 'Count'; 'Country'; 'State'; 'City'; 'Zip Code'; 'Lat Long'; 'Latitude'; and 'Longitude', which are used to give geographic information regarding the customer. Other attributes include 'Gender' and 'Senior Citizen', which capture demographic information; 'Partner' and 'Dependents', which capture information regarding the customer's relationships; and 'Tenure Months', which captures the amount of time the customer has subscribed to the organization. Other attributes include 'Phone Service', 'Multiple Lines', 'Internet Service', 'Online Security', 'Online Backup', 'Device Protection', 'Tech Support', 'Streaming TV', 'Streaming Movies', and others, which describe different services the company provides. The attribute 'Contract' comprises details about the kind of contract the customer has, while 'Paperless Billing' and 'Payment Method' represent the mode of payment. The attributes 'Monthly Charges' and 'Total Charges' are the money-related attributes of the customer's subscription. The dataset also contains the 'Churn Label', 'Churn Value', 'Churn Score', 'CLTV', and 'Churn Reason' columns specific to churn prediction. Most importantly, this dataset exhibits class imbalance in its nature: 26.54% of the data represent churn customers, whereas 73.46% represent non-churn customers. This class imbalance poses a challenge in predictive modeling and hence needs the use of suitable techniques to handle it properly. The data ratio is illustrated in Fig. 3.3.1.

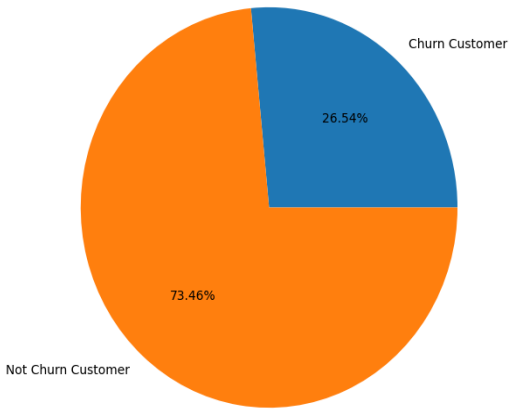

Fig: 3.3.1. Customer Data Ratio

3.4 Proposed Methodology

As we continue our analysis, it has become evident that refining and optimizing the dataset will be crucial for achieving accurate and meaningful results. We believe that by implementing these measures, we can significantly enhance the quality and reliability of our analysis.

Data Refinement: Our first step will be to clean the dataset—eliminate columns that are not useful, are duplicated, are highly correlated, or display low variance in the columns. This process will help in slimming down our dataset, as it will reduce it to the most meaningful features needed for our analysis. More specifically, we will exclude the columns CustomerID, City, Zip Code, Country, State, and Internet Service, which may not provide crucial insights for the specific purposes of our research. Furthermore, we will check for duplicate information in the dataset, with the possibility of dropping Latitude and Longitude, if and when necessary.

Data Transformation: Next, we need to standardize and make the data machine learning algorithm-ready. That will be done for all numerical columns—Tenure Months, Monthly Charges, Total Charges, and CLTV (Customer Lifetime Value)—to one standard scale by using MinMaxScaler. In addition, we will perform categorical conversion, or one-hot encoding, for features like Contract and Payment Method to present them more effectively in our analysis. Further, we simplify some of the categorical variables so that they can be represented more easily in a Boolean format, thereby making it easier for our machine learning models to interpret them.

Handling Mismatched Data Types: We will also deal with data type mismatches such as changing the Total Charges column from being read as a string to being treated as numerical data. We will also handle specific values, such as 'No phone service' and 'No internet service', being treated as 'No', to make sure that our data is consistent.

Regularization for Model Stability: Regularization with the Elastic Net technique will be applied to enhance the stability and generalization performance of the models. This method leverages the advantages of the Lasso and Ridge regularization procedures while mitigating overfitting risks.

Addressing Class Imbalance: Ensuring balance in our dataset is crucial for fair and accurate predictive modeling. To address class imbalance, we'll utilize the SMOTE technique, which generates synthetic samples for minority classes. The data ratio after applying SMOTE technique is illustrated in Fig. 3.4.1.

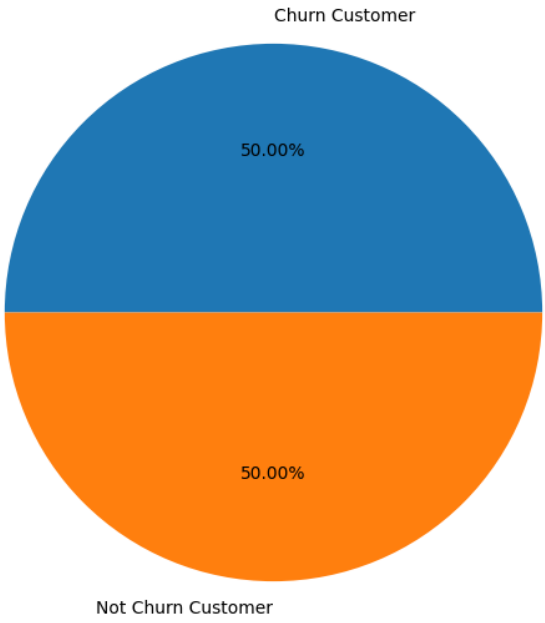

Fig: 3.4.1. Customer Data Ratio After Data Augmentation

We'll split our dataset into 70% for training and 30% for testing, applying SMOTE exclusively to the training data to maintain fairness in our analysis.

Hyperparameter Tuning: We will optimize our models to be the best through hyperparameter tuning via GridSearchCV. Through this systematic approach, one can determine the best model configuration from multiple candidates, thus making sure it delivers the best performance for the given research objectives.

Evaluation Metrics: After the parameters have been fine-tuned, we will be ready to assess our models using different types of metrics, for example, accuracy, precision, recall, F1-score, or ROC AUC. Such assessment will be illustrative of model performance in different variations of parameters and will allow making decisions based on their effectiveness.

3.5 Proposed Model Workflow

To effectively predict customer churn, a comprehensive series of actions is required to ensure accuracy and reliability. The block diagram in Fig. 3.5.1 outlines the proposed model workflow for this study, detailing each critical step in the process. This workflow includes data preprocessing, feature extraction, data augmentation, and the application of Grid Search Cross-Validation on various machine learning models.

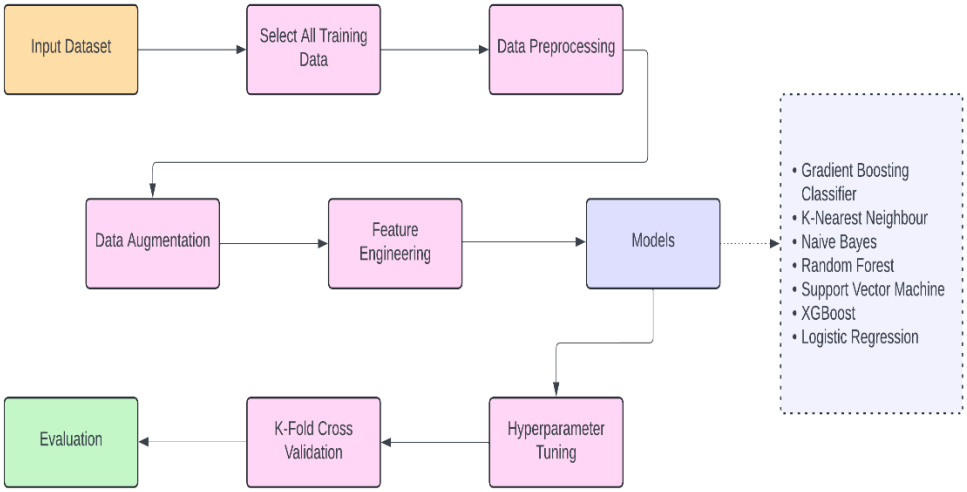

Fig. 3.5.1. Proposed Model Workflow

3.6 Implementation Requirements

There are some requirements to implemented this project.

Hardware or Software Requirements:

- Windows 10 or above operating system
- Portable or in-built Hard Disk above 300 GB
- Minimum 4 GB RAM
- Google Chrome or Mozilla Firefox

Developing Tools:

- Python
- Jupiter

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Introduction

In this chapter, we present and analyze the experimental results obtained from applying various machine learning techniques to predict customer churn in the telecom industry. By employing our proposed framework, we have been able to thoroughly test and validate our approach.

4.2 Experimental Results

In this section, we will demonstrate the experiments we conducted to find the optimal hyperparameters using Grid Search Cross-Validation for various machine learning techniques. We performed hyperparameter tuning of these techniques using Grid Search Cross-Validation with K-fold sizes of 5 and 10.

The results of hyperparameter tuning for each technique with a K-fold size of 5 are illustrated in the following figures.

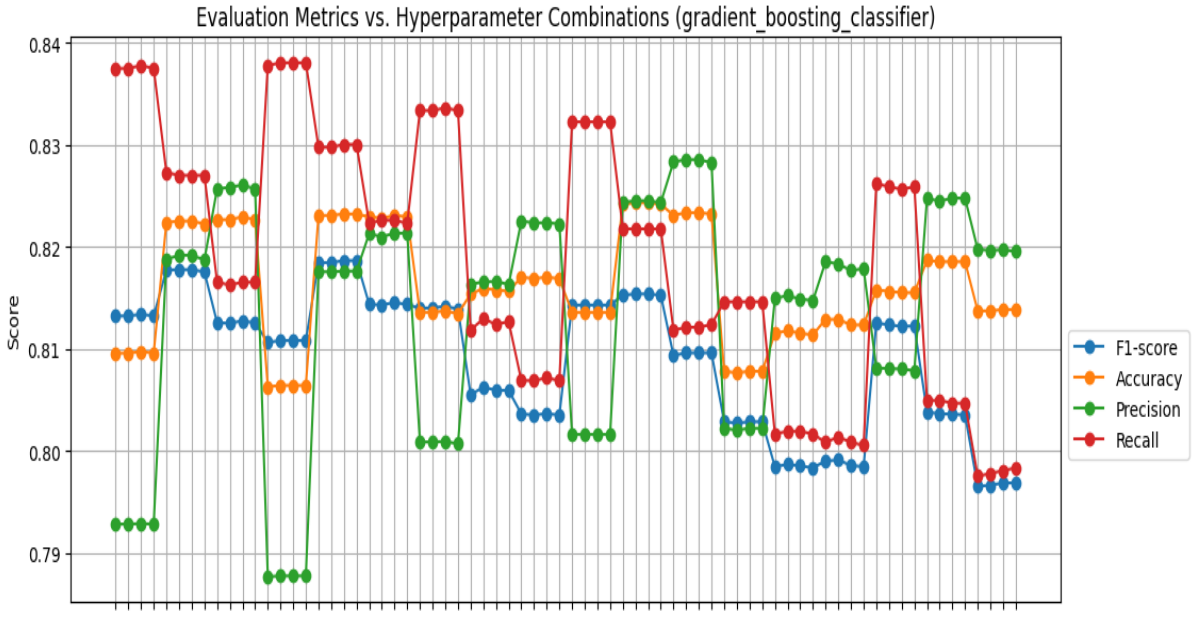

Fig 4.2.1. Hyperparameter Combinations of Gradient Boosting Classifier (K = 5)

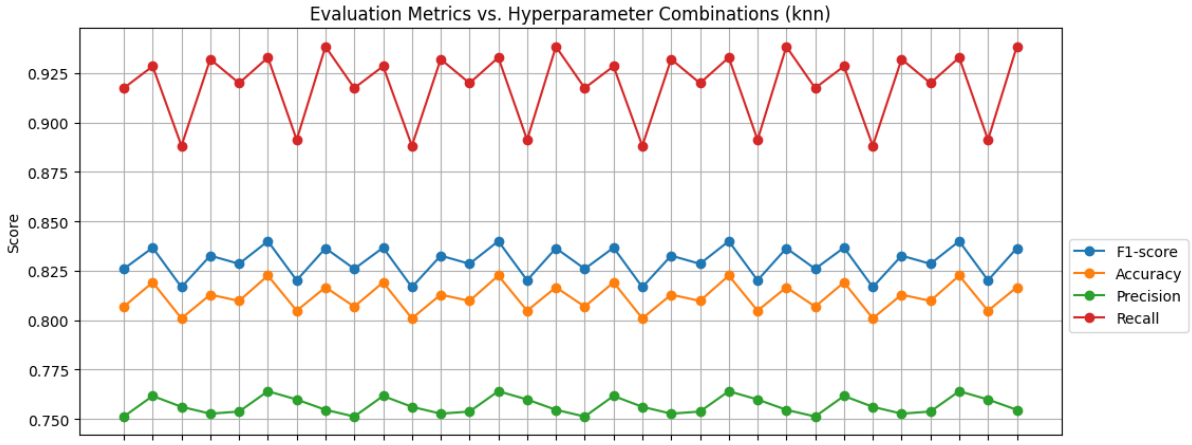

Fig 4.2.2. Hyperparameter Combinations of K-Nearest Neighbors (K = 5)

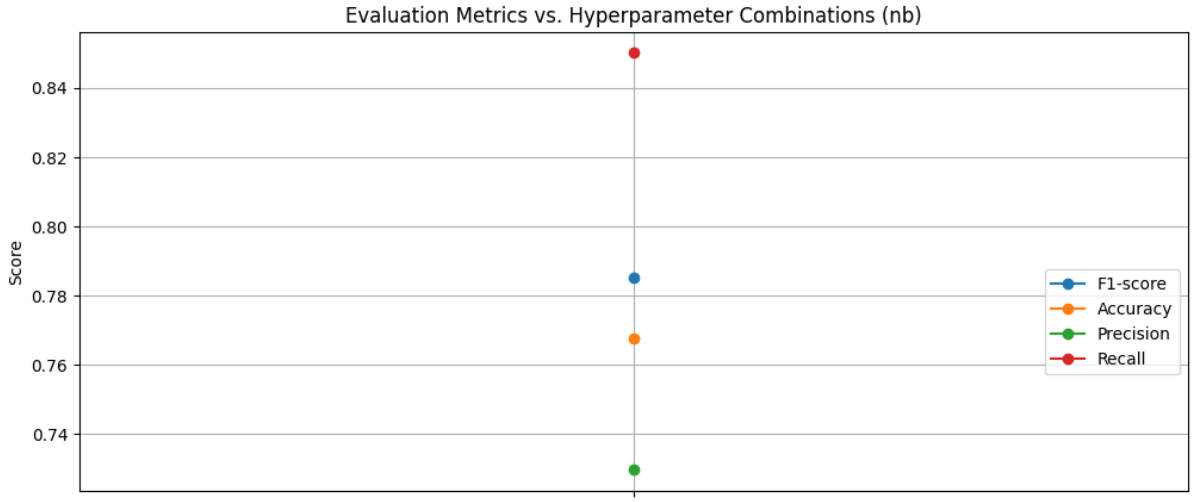

Fig 4.2.3. Hyperparameter Combinations of Naïve Bayes (K = 5)

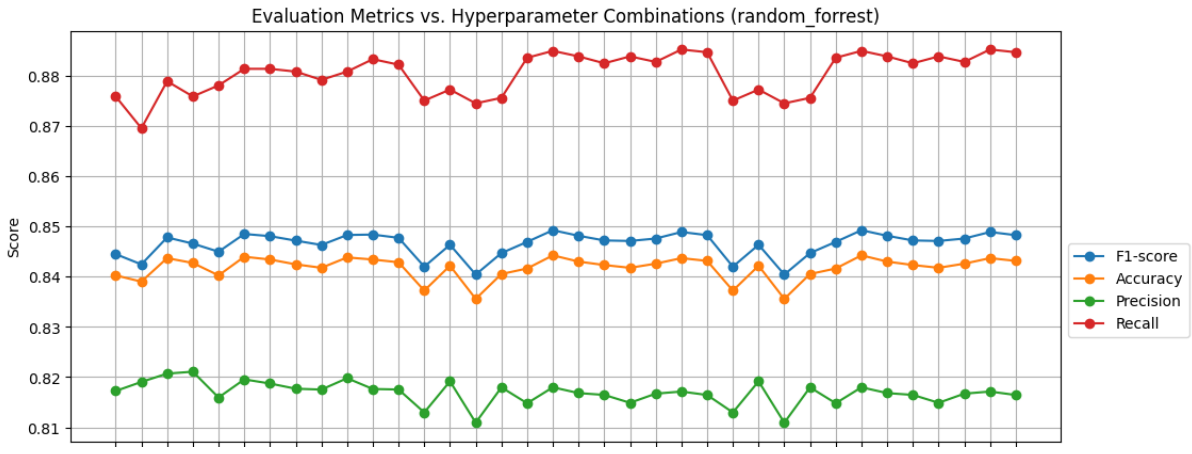

Fig 4.2.4. Hyperparameter Combinations of Random Forest (K = 5)

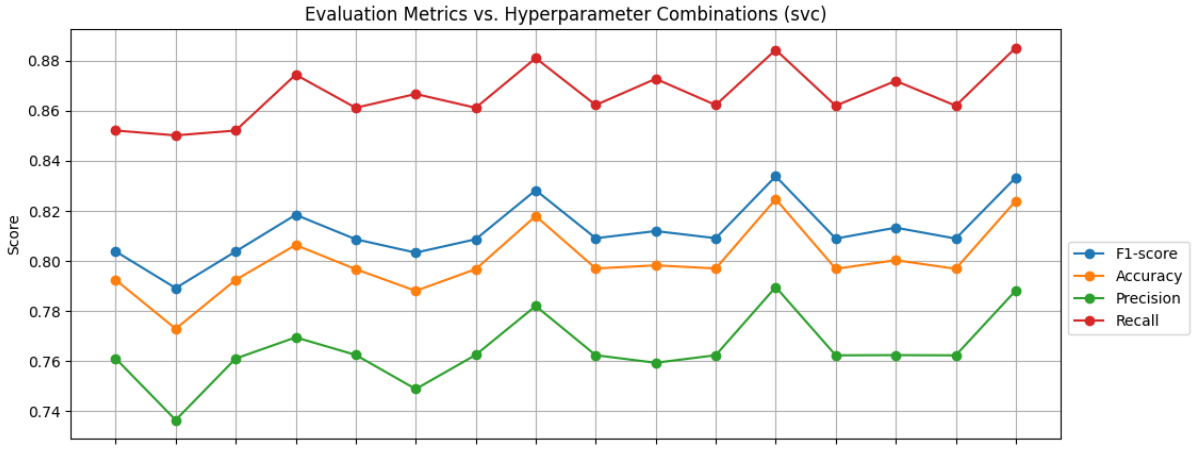

Fig 4.2.5. Hyperparameter Combinations of Support Vector Machine (K = 5)

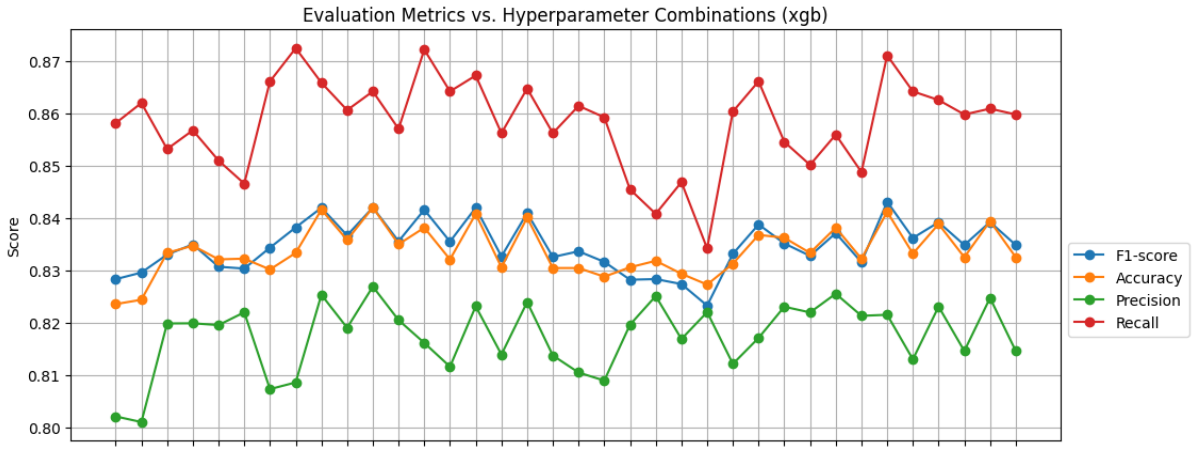

Fig 4.2.6. Hyperparameter Combinations of XGBoost (K = 5)

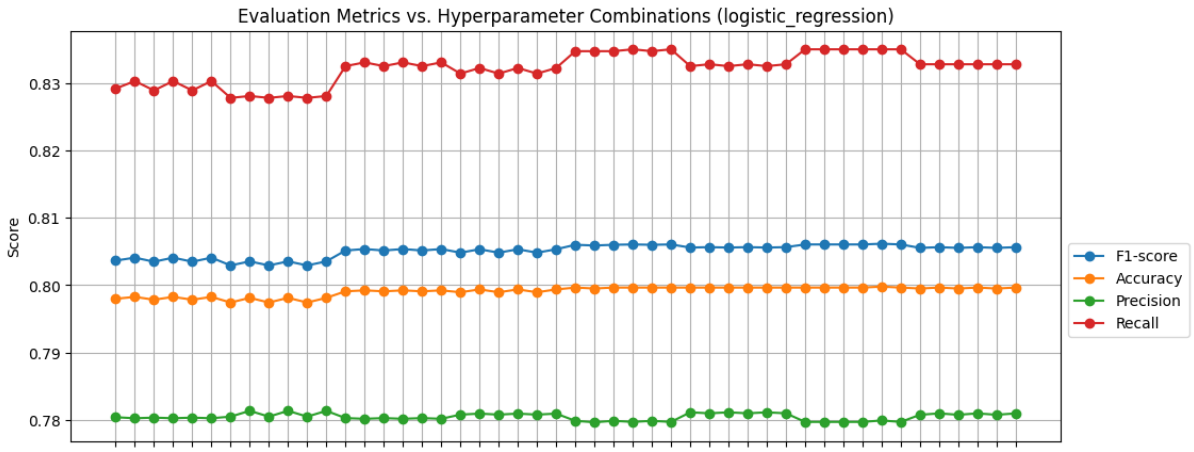

Fig 4.2.7. Hyperparameter Combinations of Logistic Regression (K = 5)

For K = 5, The Fig 4.2.1, Fig 4.2.2, Fig 4.2.3, Fig 4.2.4, Fig 4.2.5, Fig 4.2.6, and Fig 4.2.7 illustrates the results of hyperparameter tuning for optimizing the performance of the models. It shows the variation in performance metrics as different values of the

hyperparameter are tested. Each point on the graph represents a specific configuration of the hyperparameter, with performance metrics such as accuracy, F1 score, precision, and recall plotted on the y-axis against the corresponding values of the hyperparameter on the x-axis.

The results of hyperparameter tuning for each technique with a K-fold size of 10 are illustrated in the following figures.

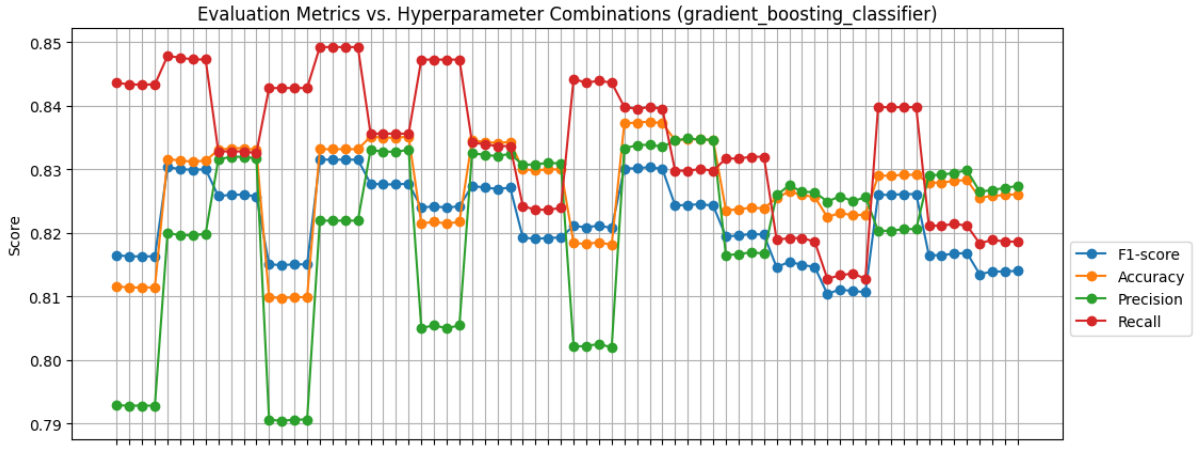

Fig 4.2.8. Hyperparameter Combinations of Gradient Boosting Classifier (K = 10)

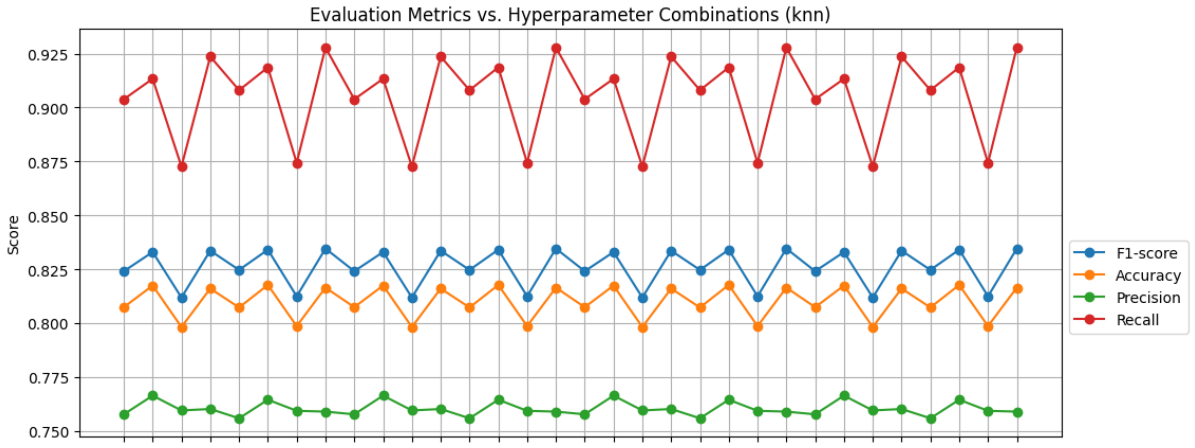

Fig 4.2.9. Hyperparameter Combinations of K-Nearest Neighbors (K = 10)

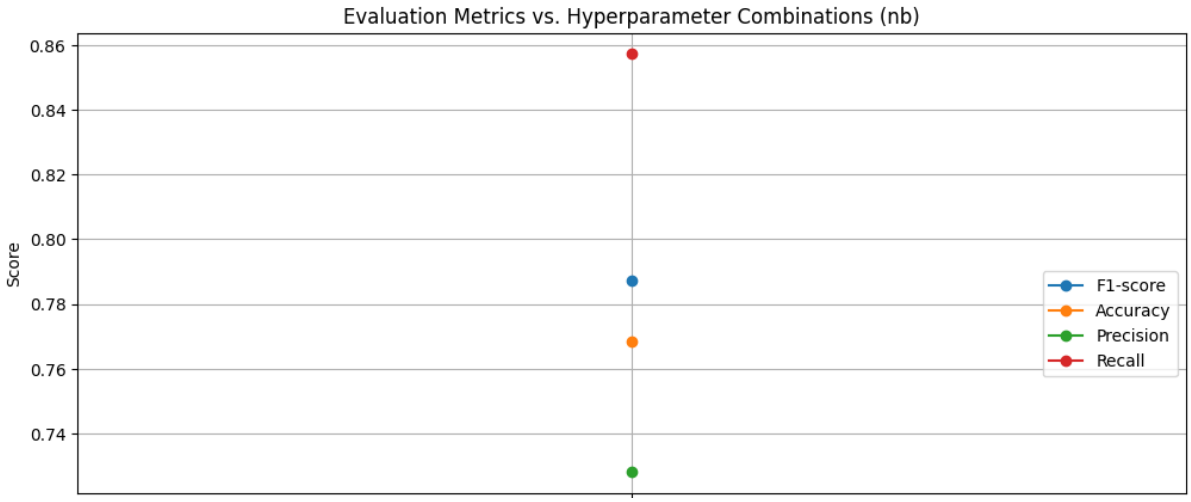

Fig 4.2.10. Hyperparameter Combinations of Naïve Bayes (K = 10)

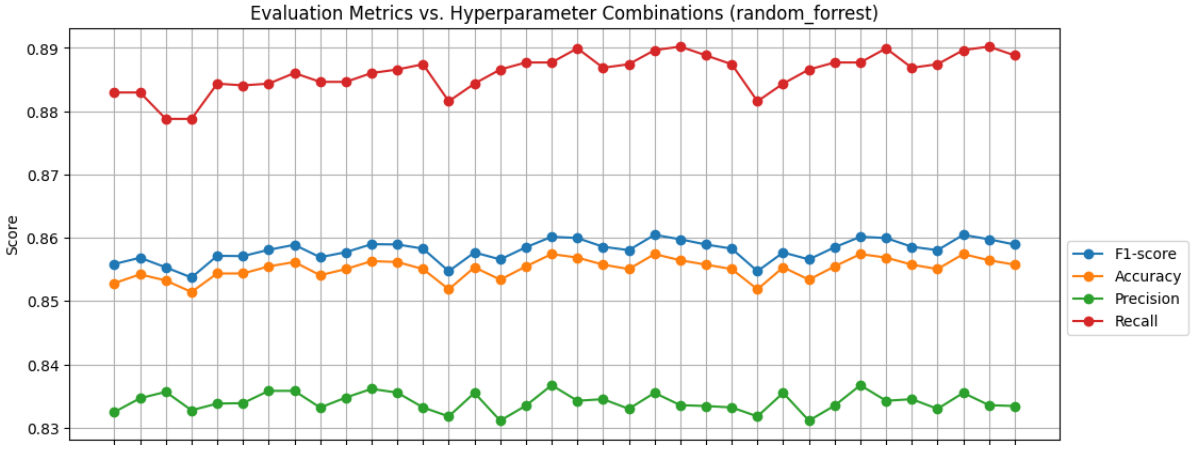

Fig 4.2.11. Hyperparameter Combinations of Random Forest (K = 10)

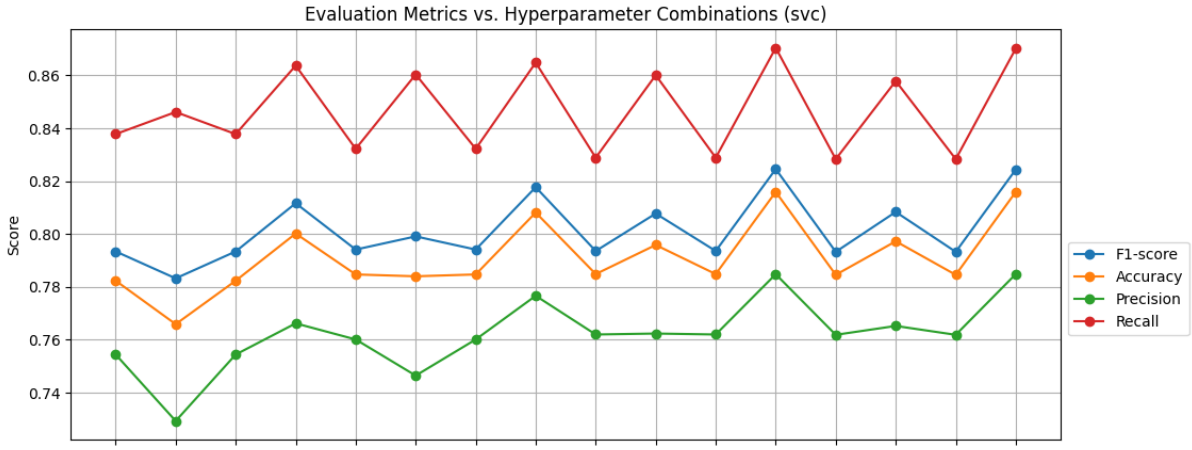

Fig 4.2.12. Hyperparameter Combinations of Support Vector Machine (K = 10)

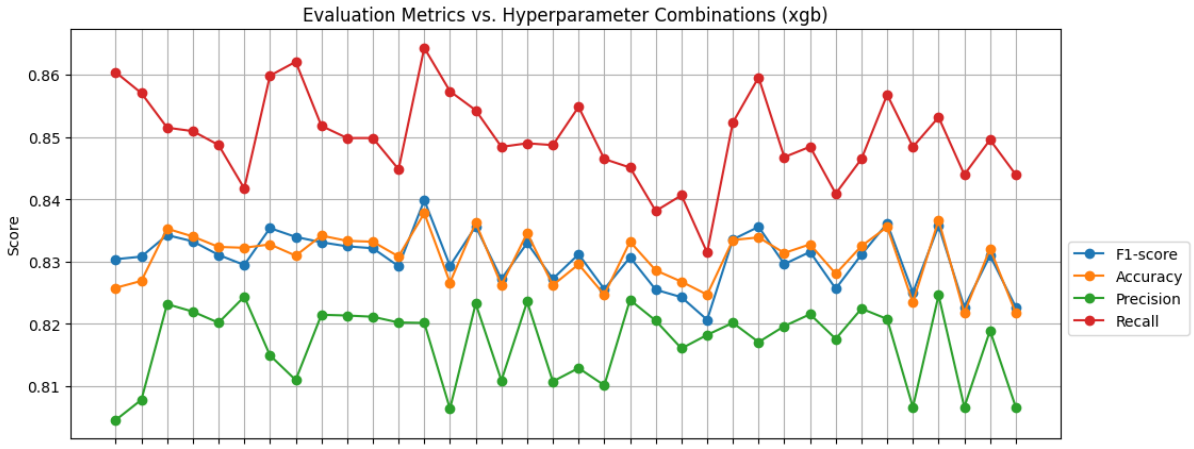

Fig 4.2.13. Hyperparameter Combinations of XGBoost (K = 10)

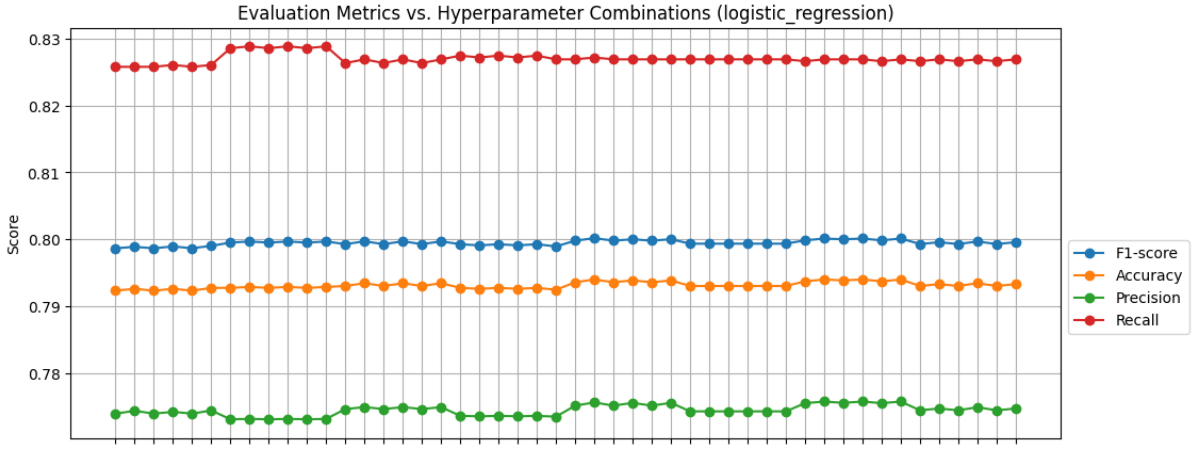

Fig 4.2.14. Hyperparameter Combinations of Logistic Regression (K = 10)

For K = 10, The Fig 4.2.8, Fig 4.2.9, Fig 4.2.10, Fig 4.2.11, Fig 4.2.12, Fig 4.2.13, and Fig 4.2.14 illustrates the results of hyperparameter tuning for optimizing the performance of the models. It shows the variation in performance metrics as different values of the hyperparameter are tested. Each point on the graph represents a specific configuration of the hyperparameter, with performance metrics such as accuracy, F1 score, precision, and recall plotted on the y-axis against the corresponding values of the hyperparameter on the x-axis.

4.3 Discussion

For each technique, we report the best performance metrics, including accuracy, precision, F1 score, and recall. These metrics are depicted in the following diagrams, providing a comprehensive view of how each model performed with the best hyperparameters.

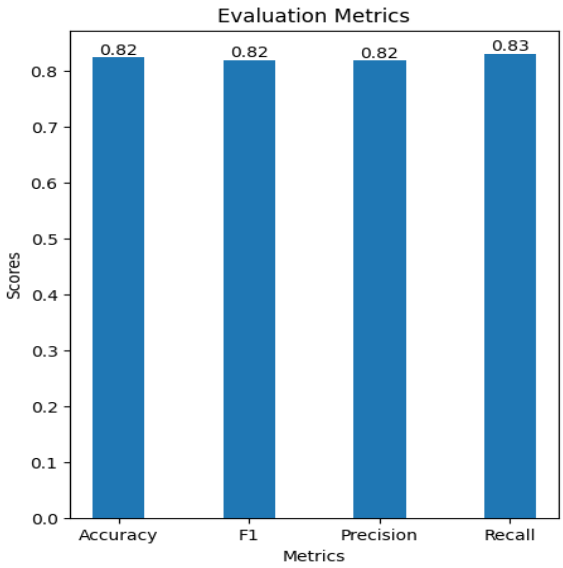

Fig 4.3.1. Evaluation Metrics of Gradient Boosting Classifier (K = 5)

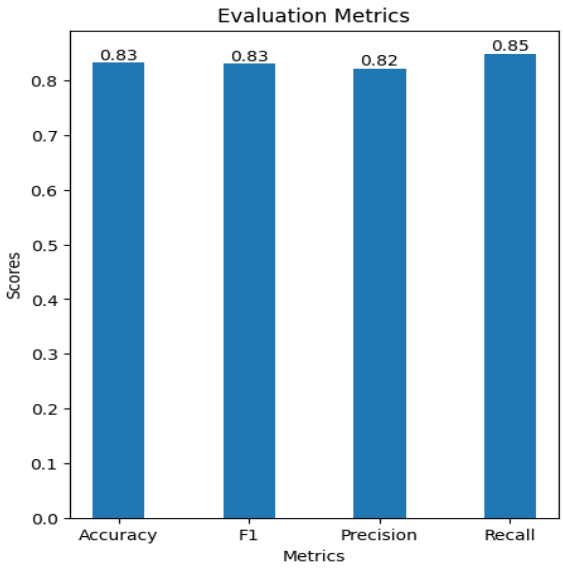

Fig 4.3.2. Evaluation Metrics of Gradient Boosting Classifier (K = 10)

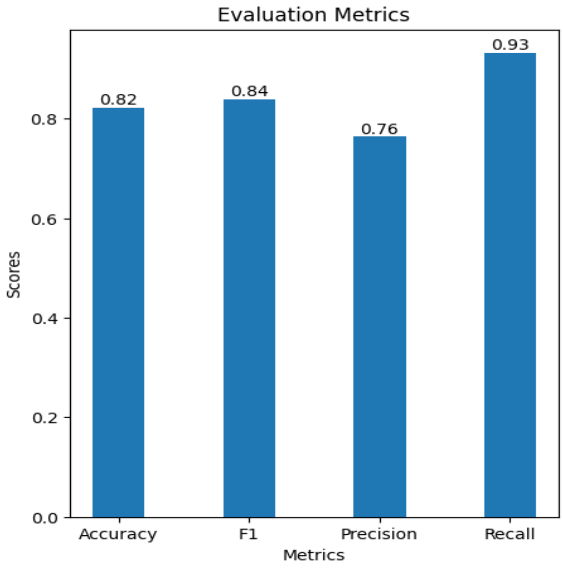

Fig 4.3.3. Evaluation Metrics of K-Nearest Neighbors (K = 5)

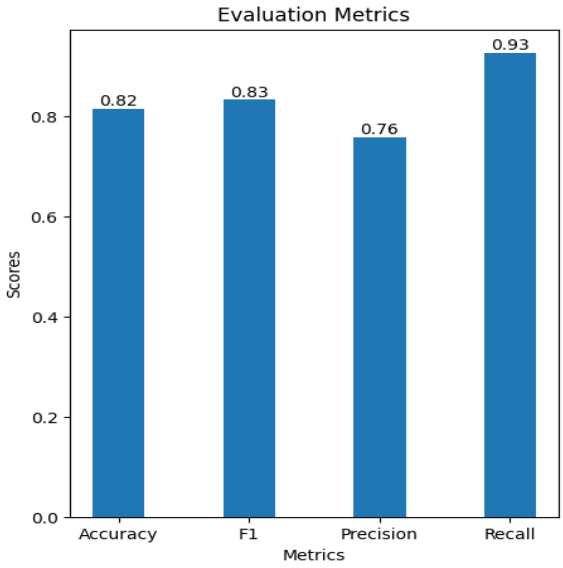

Fig 4.3.4. Evaluation Metrics of K-Nearest Neighbors (K = 10)

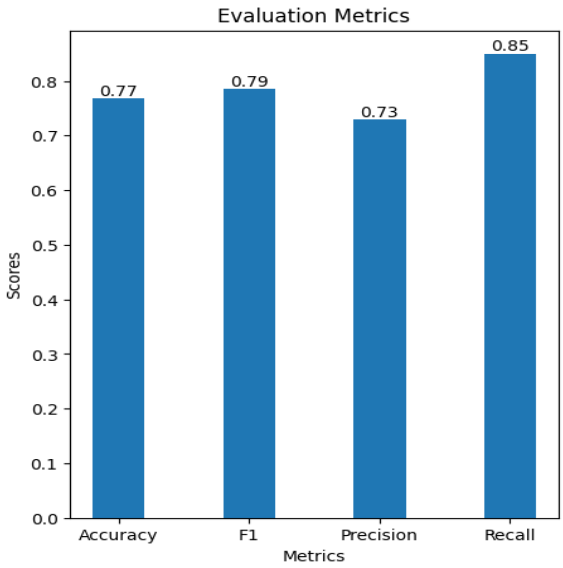

Fig 4.3.5. Evaluation Metrics of Naïve Bayes (K = 5)

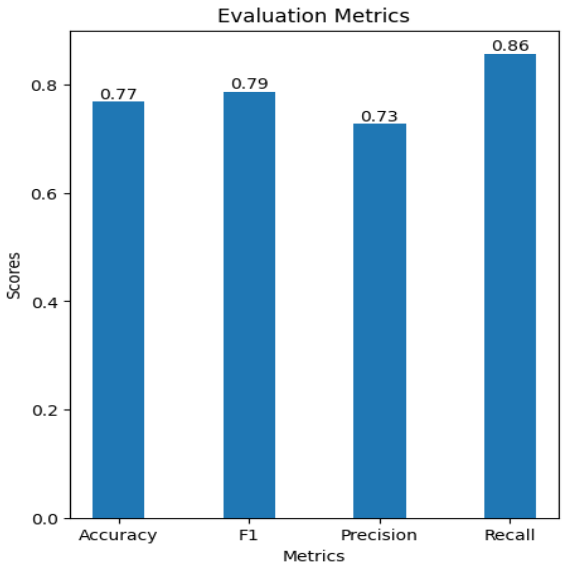

Fig 4.3.6. Evaluation Metrics of Naïve Bayes (K = 10)

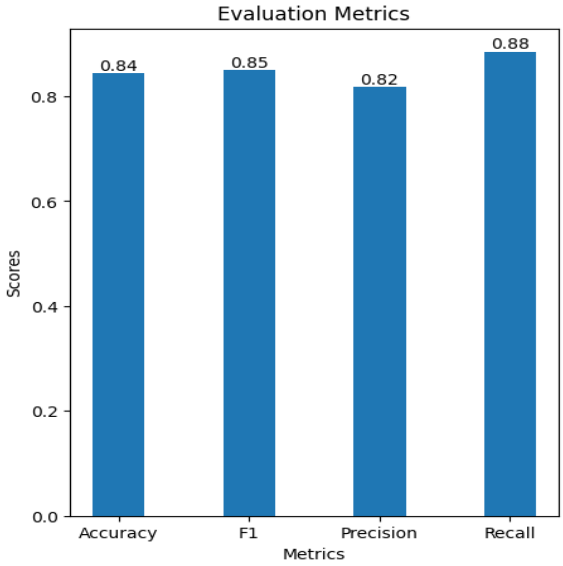

Fig 4.3.7. Evaluation Metrics of Random Forest (K = 5)

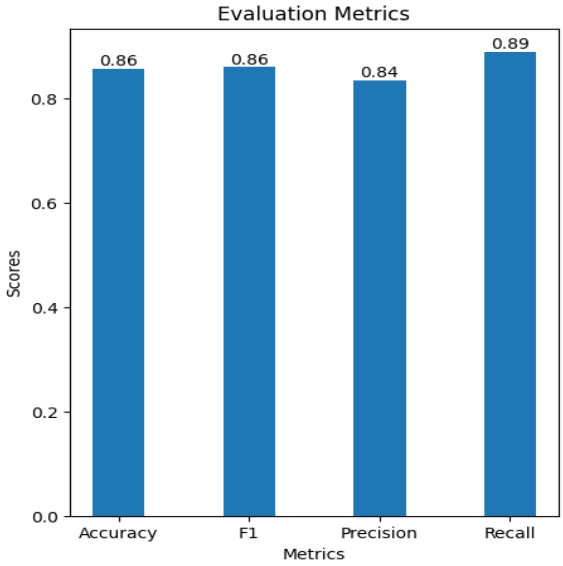

Fig 4.3.8. Evaluation Metrics of Random Forest (K = 10)

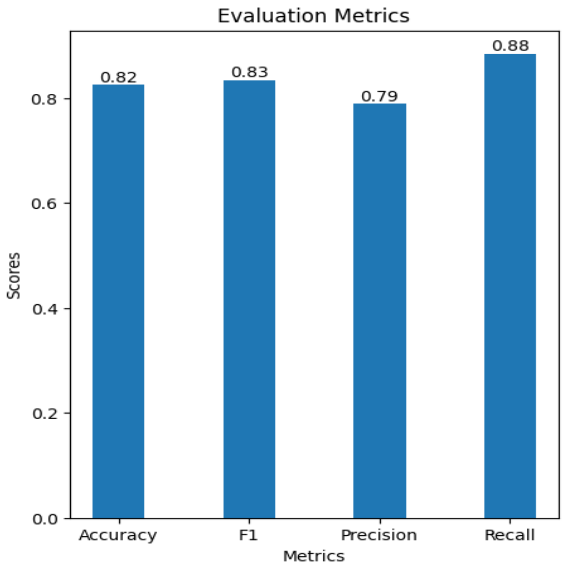

Fig 4.3.9. Evaluation Metrics of Support Vector Machine (K = 5)

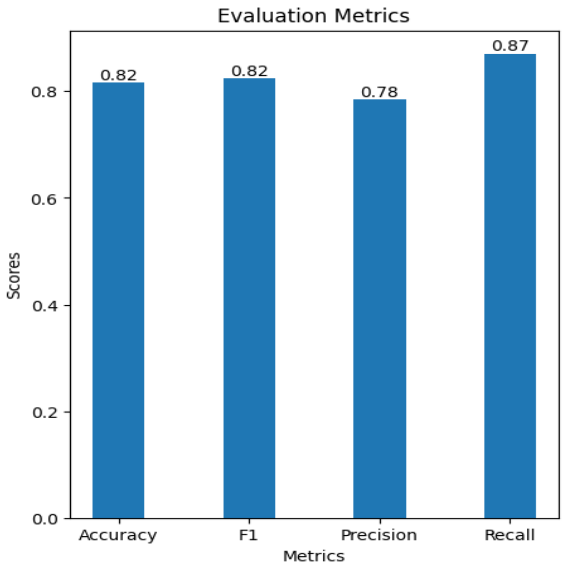

Fig 4.3.10. Evaluation Metrics of Support Vector Machine (K = 10)

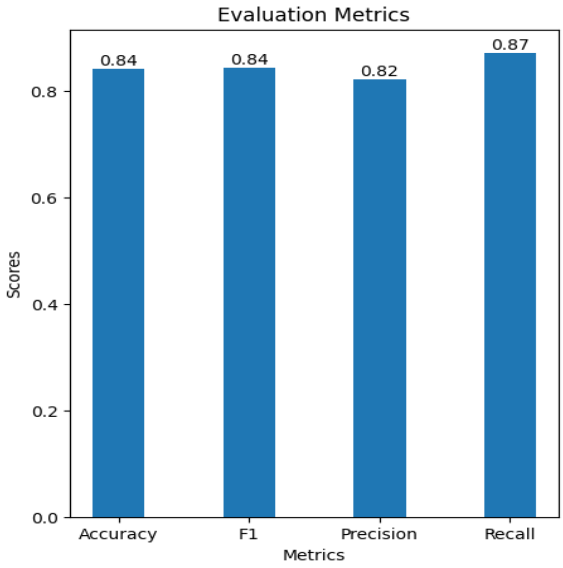

Fig 4.3.11. Evaluation Metrics of XGBoost (K = 5)

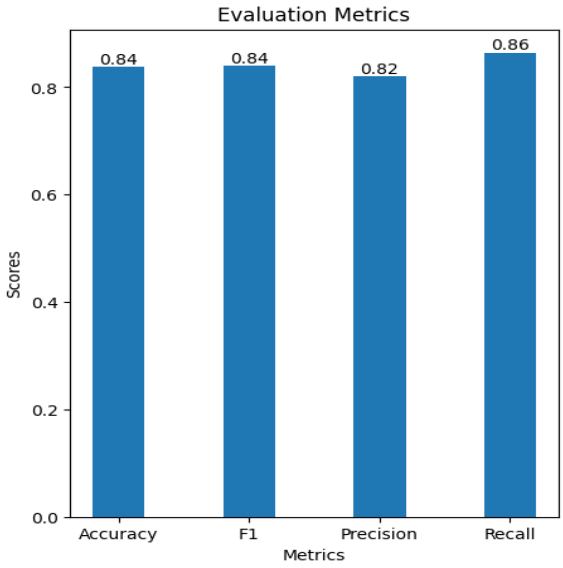

Fig 4.3.12. Evaluation Metrics of XGBoost (K = 10)

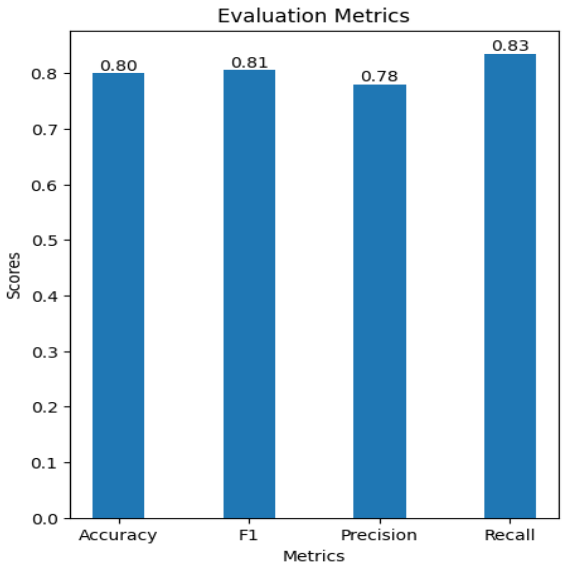

Fig 4.3.13. Evaluation Metrics of Logistic Regression (K = 5)

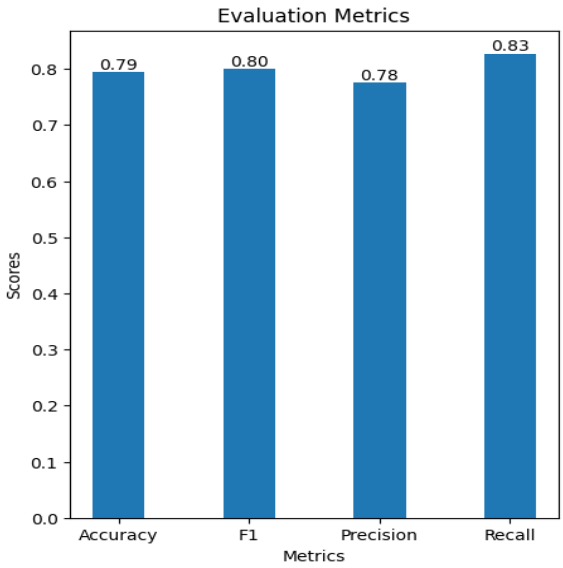

Fig 4.3.14. Evaluation Metrics of Logistic Regression (K = 10)

The Fig 4.3.1 and Fig 4.3.2 illustrates the performance evaluation metrics of the Gradient Boosting Classifier for K = 5 and K = 10. For K = 5, the performance metrics are as follows: accuracy is 82%, F1 score is 82%, precision is 82%, and recall is 83%. This uniformity in accuracy, F1 score, and precision indicates that the model performs consistently across these evaluation criteria, while the slightly higher recall value of 83% suggests a good ability to identify relevant instances. For K = 10, the performance metrics show slight improvements: accuracy and F1 score both increased to 83%, precision remained at 82%, and recall improved further to 85%. The improvement in recall indicates the model's enhanced ability to identify relevant instances with a higher value of K, while maintaining consistent precision. The increase in the F1 score and accuracy suggests that the overall predictive performance of the model has slightly improved with the higher K value.

The Fig 4.3.3 and Fig 4.3.4 illustrates the performance evaluation metrics of the K-Nearest Neighbors for K = 5 and K = 10. For K = 5, the performance metrics are as follows: accuracy is 82%, F1 score is 84%, precision is 76%, and recall is 93%. This indicates that while the model has high recall, identifying 93% of relevant instances, its precision is lower at 76%, meaning that there are more false positives. The high F1 score of 84% reflects a balance between precision and recall, weighted more towards recall. For K = 10, the performance metrics are consistent: accuracy is 82%, F1 score is 83%, precision is 76%,

and recall is 93%. This shows that increasing K to 10 maintains the same accuracy and recall while slightly decreasing the F1 score from 84% to 83%. The precision remains the same at 76%, indicating that the model's ability to correctly identify positive instances has not changed with an increase in K.

The Fig 4.3.5 and Fig 4.3.6 illustrates the performance evaluation metrics of the Naïve Bayes for K = 5 and K = 10. For K = 5, the performance metrics are as follows: accuracy is 77%, F1 score is 79%, precision is 73%, and recall is 85%. This indicates that while the model has a high recall, identifying 85% of relevant instances, its precision is lower at 73%, meaning that there are more false positives. The F1 score of 79% reflects a balance between precision and recall, with a slight emphasis on recall. For K = 10, the performance metrics are as follows: accuracy remains the same at 77%, F1 score is 79%, precision is 73%, and recall improves slightly to 86%. This shows that increasing K to 10 maintains the same accuracy and precision while enhancing recall to 86%, indicating a better ability to identify relevant instances. The consistent F1 score of 79% reflects a balanced performance between precision and recall at this higher K value.

The Fig 4.3.7 and Fig 4.3.8 illustrates the performance evaluation metrics of the Random Forest for K = 5 and K = 10. For K = 5, the performance metrics are as follows: accuracy is 84%, F1 score is 85%, precision is 82%, and recall is 88%. This indicates a strong performance, with the model accurately identifying 88% of relevant instances (high recall) and maintaining a precision of 82%, which reflects a good balance between correctly identifying positive instances and minimizing false positives. The F1 score of 85% demonstrates an overall balanced performance between precision and recall. For K = 10, the performance metrics improve slightly: accuracy increases to 86%, F1 score rises to 86%, precision improves to 84%, and recall increases to 89%. These results indicate that increasing K to 10 enhances the model's ability to correctly identify relevant instances (higher recall) and correctly classify positive instances (higher precision). The increase in accuracy to 86% shows that the overall predictive performance of the model has improved. The consistent F1 score of 86% reflects a well-maintained balance between precision and recall at the higher K value.

The Fig 4.3.9 and Fig 4.3.10 illustrates the performance evaluation metrics of the Support Vector Machine for K = 5 and K = 10. For K = 5, the performance metrics are as follows:

accuracy is 82%, F1 score is 83%, precision is 79%, and recall is 88%. This indicates a strong performance, with the model accurately identifying 88% of relevant instances (high recall) and maintaining a precision of 79%, which reflects a balance between correctly identifying positive instances and minimizing false positives. The F1 score of 83% demonstrates an overall balanced performance between precision and recall. For K = 10, the performance metrics are as follows: accuracy remains the same at 82%, F1 score also remains the same at 82%, precision decreases slightly to 78%, and recall decreases slightly to 87%. These results indicate that increasing K to 10 maintains a similar level of accuracy and F1 score while showing slight decreases in precision and recall. The F1 score of 82% reflects a well-maintained balance between precision and recall at the higher K value.

The Fig 4.3.11 and 4.3.12 illustrates the performance evaluation metrics of the XGBoost for K = 5 and K = 10. For K = 5, the performance metrics are as follows: accuracy is 84%, F1 score is 84%, precision is 82%, and recall is 87%. This indicates a strong performance, with the model accurately identifying 87% of relevant instances (high recall) and maintaining a precision of 82%, which reflects a good balance between correctly identifying positive instances and minimizing false positives. The F1 score of 84% demonstrates an overall balanced performance between precision and recall. For K = 10, the performance metrics are as follows: accuracy remains the same at 84%, F1 score also remains the same at 84%, precision remains at 82%, and recall decreases slightly to 86%. These results indicate that increasing K to 10 maintains a similar level of accuracy and F1 score while showing a slight decrease in recall. The F1 score of 84% reflects a well-maintained balance between precision and recall at the higher K value.

The Fig 4.3.13 and 4.3.14 illustrates the performance evaluation metrics of the Logistic Regression for K = 5 and K = 10. For K = 5, the performance metrics are as follows: accuracy is 80%, F1 score is 81%, precision is 78%, and recall is 83%. This indicates a solid performance, with the model accurately identifying 83% of relevant instances (recall) and maintaining a precision of 78%, which reflects a balance between correctly identifying positive instances and minimizing false positives. The F1 score of 81% demonstrates an overall balanced performance between precision and recall. For K = 10, the performance metrics are as follows: accuracy is 79%, F1 score is 80%, precision remains at 78%, and recall also remains at 83%. These results indicate that increasing K to 10 results in a slight decrease in accuracy and F1 score, while precision and recall remain consistent. The F1

score of 80% reflects a well-maintained balance between precision and recall at the higher K value.

Based on the results from the K-fold size of 5, the maximum accuracy achieved was 84% by both Random Forest and XGBoost, while the lowest accuracy was exhibited by Naive Bayes at only 77%. All other techniques showed average performance, with accuracies ranging around 82% and encompassing the Gradient Boosting Classifier, KNN, Logistic Regression, and SVM. For K-fold size 10, the top performer remained to be Random Forest, this time achieving an accuracy of 86%. On the other hand, Naive Bayes proved not effective again, with an accuracy of 77%. The rest of the techniques presented a disparate performance, with accuracies ranging from 80% to 84%.

These results demonstrate how Random Forest and XGBoost are competitive strategies for predicting customer churn, with a clear relative position in the strengths and weaknesses of other machine learning models. Therefore, the detailed analysis and comparison described in this research will drive future research in selecting and optimizing machine learning techniques for customer churn prediction in the telecom sector.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Summary of the study

This study tackles the critical issue of customer churn in the telecommunications industry. Customer churn is about customers cutting their service subscriptions. It significantly affects the profitability of telecom companies. Effective churn prediction helps companies put into place proactive retention strategies. With the help of the IBM Telco dataset, which contains 7,044 customer records, and the churn rate is 26.54%, the study evaluates the performance of seven machine learning techniques. Grid Search Cross Validation is used with K-Fold sizes of 5 and 10 to fine-tune the hyperparameters and increase the model accuracy.

Furthermore, this paper incorporates comprehensive data preprocessing, such as cleaning and feature selection, to handle class imbalance and increase model performance. Experimental results reveal that Random Forest and XGBoost yield the highest accuracy rates of 84% with 5-fold and 86% with 10-fold cross-validation. Naive Bayes performs the lowest, with an accuracy of 77%. The major challenges are class balancing, avoidance of overfitting, and dealing with computation demands of model training. However, the study determines the most effective machine learning algorithms for predicting customer churn and provides useful insight for telecom companies to retain customers and minimize churn rates.

5.2 Conclusions

This thesis successfully presents the critical issue of customer churn in the telecommunications industry by evaluating and comparing the effectiveness of various machine learning models. Using the IBM Telco dataset and advanced techniques like K-Fold Cross Validation and Grid Search for hyperparameter optimization, the study analyzes seven machine learning algorithms: Gradient Boosting Classifier, K-Nearest Neighbors, Logistic Regression, Naive Bayes, Random Forest, Support Vector Machine, and XGBoost.

The findings of this research show that Random Forest and XGBoost models yield the highest accuracy of all models, while Naive Bayes is the least effective. The research therefore identifies data preprocessing, important feature selection, and appropriate handling of class imbalance as ways to enhance model performance. By extensive experimentation and analysis, the study gives important insights and a stronger framework for telecom organizations to predict and mitigate customer churn.

This thesis contributes to the field by providing the knowledge of more suitable machine learning models in predicting churn, which can be used by the telecom companies to retain customers and improve their profitability. The methodologies and results presented here may serve as a foundation for further research in this area and its application in practice for the preparation of customer retention strategies.

5.3 Implication for further study

This thesis lays a strong foundation for future research in the domain of customer churn prediction. In future we will –

- Validate models with datasets from different telecom companies and industries.
- Explore sophisticated feature engineering methods for enhanced model performance.
- Investigate neural networks and other deep learning techniques.
- Develop systems for continuous learning and adaptation to new data.
- Compare current models with new algorithms to improve churn prediction.

REFERENCE

[1] V. L. Miguéis, D. Van den Poel, A. S. Camanho, and J. Falcão e Cunha, "Modeling partial customer churn: On the value of first product-category purchase sequences," Expert Systems with Applications, vol. 39, no. 12, pp. 11250–11256, Sep. 2012, doi: https://doi.org/10.1016/j.eswa.2012.03.073.

[2] H. K. Thakkar, A. Desai, S. Ghosh, P. Singh, and G. Sharma, "Clairvoyant: AdaBoost with Cost-Enabled Cost-Sensitive Classifier for Customer Churn Prediction," Computational Intelligence and Neuroscience, vol. 2022, pp. 1–11, Jan. 2022, doi: https://doi.org/10.1155/2022/9028580.

[3] X. Miao and H. Wang, "Customer Churn Prediction on Credit Card Services using Random Forest Method," Proceedings of the 2022 7th International Conference on Financial Innovation and Economic Development (ICFIED 2022), 2022, doi: https://doi.org/10.2991/aebmr.k.220307.104.

[4] A. Mishra and U. S. Reddy, "A comparative study of customer churn prediction in telecom industry using ensemble based classifiers," IEEE Xplore, Nov. 01, 2017. doi: https://doi.org/10.1109/ICICI.2017.8365230

[5] M. Bogaert and L. Delaere, "Ensemble Methods in Customer Churn Prediction: A Comparative Analysis of the State-of-the-Art," Mathematics, vol. 11, no. 5, p. 1137, Feb. 2023, doi: https://doi.org/10.3390/math11051137.

[6] R. Sudharsan and E. N. Ganesh, "A Swish RNN based customer churn prediction for the telecom industry with a novel feature selection strategy," Connection Science, vol. 34, no. 1, pp. 1855–1876, Jun. 2022, doi: https://doi.org/10.1080/09540091.2022.2083584.

[7] X. Xiahou and Y. Harada, "B2C E-Commerce Customer Churn Prediction Based on K-Means and SVM," Journal of Theoretical and Applied Electronic Commerce Research, vol. 17, no. 2, pp. 458–475, Apr. 2022, doi: https://doi.org/10.3390/jtaer17020024.

[8] Sofi, "How to Build a Churn Prediction Model to Predict Customer Churn," Thoughts about Product Adoption, User Onboarding and Good UX | Userpilot Blog, Sep. 19, 2023. https://userpilot.com/blog/churn-prediction/

[9] P. Lalwani, M. K. Mishra, J. S. Chadha, and P. Sethi, "Customer churn prediction system: a machine learning approach," Computing, vol. 104, Feb. 2021, doi: https://doi.org/10.1007/s00607-021-00908-y.

[10] B. Prabadevi, R. Shalini, and B. R. Kavitha, "Customer churning analysis using machine learning algorithms," International Journal of Intelligent Networks, vol. 4, Jun. 2023, doi: https://doi.org/10.1016/j.ijin.2023.05.005.

[11] "Customer Churn Prediction in Telecommunication Industry Using Deep Learning," Information Sciences Letters, vol. 11, no. 1, pp. 185–198, Jan. 2022, doi: https://doi.org/10.18576/isl/110120

[12] M. Lalwani and B. Modi, "The design and implementation of voltage stabiliser for nullifying temperature impact on PV array with grid-connected systems," International Journal of Renewable Energy Technology, vol. 13, no. 2, p. 1, 2022, doi: https://doi.org/10.1504/ijret.2022.10044159.

[13] T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis, and K. Ch. Chatzisavvas, "A comparison of machine learning techniques for customer churn prediction," Simulation Modelling Practice and Theory, vol. 55, no. 55, pp. 1–9, Jun. 2015, doi: https://doi.org/10.1016/j.simpat.2015.03.003.

[14] A. K. Ahmad, A. Jafar, and K. Aljoumaa, “Customer churn prediction in telecom using machine learning in big data platform,” Journal of Big Data, vol. 6, no. 1, Mar. 2019, doi: https://doi.org/10.1186/s40537-019-0191-6.

[15] Homeyra Khaledian, R. Saez, Jordi Vila-Valls, and X. Prats, “On the Impact of Guidance Commands Mismatch in IMM-Based Guidance Modes Identification for Aircraft Trajectory Prediction,” 2022 IEEE/AIAA 41st Digital Avionics Systems Conference (DASC), Sep. 2022, doi: https://doi.org/10.1109/dasc55683.2022.9925846.

[16] N. Heidari, P. Moradi, and A. Koochari, “An attention-based deep learning method for solving the cold-start and sparsity issues of recommender systems,” Knowledge-Based Systems, Sep. 2022, doi: https://doi.org/10.1016/j.knosys.2022.109835.

[17] S. M. Sina Mirabdolbaghi and B. Amiri, “Model Optimization Analysis of Customer Churn Prediction Using Machine Learning Algorithms with Focus on Feature Reductions,” Discrete Dynamics in Nature and Society, vol. 2022, pp. 1–20, Jun. 2022, doi: https://doi.org/10.1155/2022/5134356.

[18] N. Edwine, W. Wang, W. Song, and D. Ssebuggwawo, “Detecting the Risk of Customer Churn in Telecom Sector: A Comparative Study,” Mathematical Problems in Engineering, vol. 2022, pp. 1–16, Jul. 2022, doi: https://doi.org/10.1155/2022/8534739.

[19] T. Zhang, S. Moro, and R. F. Ramos, “A Data-Driven Approach to Improve Customer Churn Prediction Based on Telecom Customer Segmentation,” Future Internet, vol. 14, no. 3, p. 94, Mar. 2022, doi: https://doi.org/10.3390/fi14030094.

[20] R. Yahaya, O. A. Abisoye, and S. A. Bashir, “An Enhanced Bank Customers Churn Prediction Model Using A Hybrid Genetic Algorithm And K-Means Filter And Artificial Neural Network,” 2020 IEEE 2nd International Conference on Cyberspac (CYBER NIGERIA), Feb. 2021, doi: https://doi.org/10.1109/cybernigeria51635.2021.9428805.

[21] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam, and S. W. Kim, “A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector,” IEEE Access, vol. 7, pp. 60134–60149, 2019, doi: https://doi.org/10.1109/access.2019.2914999.

[22] G. Bansal and Vinay Chamola, “Lightweight Authentication Protocol for Inter Base Station Communication in Heterogeneous Networks,” Jul. 2020, doi: https://doi.org/10.1109/infocomwkshps50562.2020.9162714.

[23] Amuda, Kamorudeen A and A. B. Adeyemo, “Customers Churn Prediction in Financial Institution Using Artificial Neural Network,” Dec. 2019, doi: https://doi.org/10.48550/arxiv.1912.11346.

[24] S. Kim, K. Shin, and K. Park, “An Application of Support Vector Machines for Customer Churn Analysis: Credit Card Case,” Lecture Notes in Computer Science, pp. 636–647, 2005, doi: https://doi.org/10.1007/11539117_91.

[25] A. Keramati, H. Ghaneei, and S. M. Mirmohammadi, “Developing a prediction model for customer churn from electronic banking services using data mining,” Financial Innovation, vol. 2, no. 1, Aug. 2016, doi: https://doi.org/10.1186/s40854-016-0029-6.

[26] M. Rahman and V. Kumar, “Machine Learning Based Customer Churn Prediction In Banking,” IEEE Xplore, Nov. 01, 2020. https://ieeexplore.ieee.org/document/9297529

[27] S. Khodabandehlou and M. Zivari Rahman, “Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior,” Journal of Systems and Information Technology, vol. 19, no. 1/2, pp. 65–93, Mar. 2017, doi: https://doi.org/10.1108/jsit-10-2016-0061.

COMPARATIVE ANALYSIS OF MACHINE LEARNING TECHNIQUES FOR PREDICTING CUSTOMER CHURN IN THE TELECOM INDUSTRY

ORIGINALITY REPORT

19%	15%	9%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
2	Submitted to Daffodil International University Student Paper	2%
3	Kampa Lavanya, Juluru Jahnavi Sai Aasritha, Mohan Krishna Garnepudi, Vamsi Krishna Chellu. "Chapter 60 A Customer Churn Prediction Using CSL-Based Analysis for ML Algorithms: The Case of Telecom Sector", Springer Science and Business Media LLC, 2024 Publication	1%
4	123dok.com Internet Source	1%
5	Omer Bugra Kirgiz, Meltem Kiygi-Calli, Sendi Cagliyor, Maryam El Oraiby. "Assessing the effectiveness of OTT services, branded apps, and gamified loyalty giveaways on mobile customer churn in the telecom industry: A	1%