Comparative Evaluation of Machine Learning Models for Telecom Customer Churn Prediction Using Advanced Feature Engineering and Enhanced Ensemble Learning

Nure Jannat

Bachelor of Science

DAFFODIL INTERNATIONAL UNIVERSITY

Name: Nure Jannat
ID : 221-35-812
Title: comparative Evaluation of Machine Learning prediction using Advance Feature Engineering and Enhanced Ensemble Learning

APPROVAL

This thesis titled on "**Thesis/Project Title**", submitted by **Student Name (ID: yyy-35-yyyy)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

-- **Chairman**

Dr. Imran Mahmud
Professor & Head
Department of Software Engineering
Faculty of Science and Information Technology Daffodil International University

-- **Internal Examiner 1**

Afsana Begum
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **Internal Examiner 2**

Md. Shohel Arman
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **Internal Examiner 3**

Nadira Islam
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

-- **External Examiner**

Md Manowarul Islam
Professor
Department of Computer Science and Engineering
Jagannath University, Bangladesh

DAFFODIL INTERNATIONAL UNIVERSITY

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name	:	NURE JANNAT
Date of Birth	:	01/01/2001
Title	:	Comparative Evaluation of Machine Learning Models for Telecom Customer Churn Prediction Using Advanced Feature Engineering and Enhanced Ensemble Learning
Academic Session	:	2022-2025

I declare that this thesis is classified as:

✓ CONFIDENTIAL	(Contains confidential information under the Official Secret Act 1997)*
✓ RESTRICTED	(Contains restricted information as specified by the organization where research was done)*
✓ OPEN ACCESS	I agree that my thesis to be published as online open access (Full Text)

I acknowledge that Daffodil International University reserves the following rights:

1. The Thesis is the Property of Daffodil International University.
2. The Library of Daffodil International University has the right to make copies of the thesis for the purpose of research only.
3. The Library of Daffodil International University has the right to make copies of the thesis for academic exchange.

Certified by:

(Student's Signature)

(Supervisor's Signature)

Afsana Begum

Name of Supervisor
Date: 25/12/2025

Student ID 221-35-812
Date: 25/12/2025

THESIS DECLARATION LETTER (OPTIONAL)

Librarian,
Daffodil International University,
Daffodil Smart City,
Ashulia.Dhaka,Bangladesh

Dear Sir,

CLASSIFICATION OF THESIS AS RESTRICTED

Please be informed that the following thesis is classified as RESTRICTED for a period of three (3) years from the date of this letter. The reasons for this classification are as listed below.

Author’s Name	Nure Jannat
Thesis Title	Comparative Evaluation of Machine Learning Models for Telecom Customer Churn Prediction Using Advanced Feature Engineering and Enhanced Ensemble Learning

Reasons (i)

(ii)

(iii)

Thank you.

Yours faithfully,

____
(Supervisor’s Signature)

Date: 25/12/2025

Note: This letter should be written by the supervisor and addressed to the Librarian, *Daffodil International University* with its copy attached to the thesis.

SUPERVISOR'S DECLARATION

I hereby declare that I have checked this thesis and in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Bachelor of Science.

____

(Supervisor's Signature)

Full Name : Afsana Begum

Position : Assistant Professor & Coordinator of M.Sc

Date : 25/12/2025

STUDENT’S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Daffodil International University or any other institution.

(Student’s Signature)

Full Name : NURE JANNAT

ID Number : 221-35-812

Date : 25th December 2025

Comparative Evaluation of Machine Learning Models for Telecom Customer Churn Prediction Using Advanced Feature Engineering and Enhanced Ensemble Learning

NURE JANNAT

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Bachelor of Science

Department of Software Engineering (Major in Data Science)

DAFFODIL INTERNATIONAL UNIVERSITY

DECEMBER 2025

ACKNOWLEDGEMENTS

At first all praise is goes to Almighty Allah, the Most Gracious and the Most Merciful. He has giving me with that much of strength, patience, and wisdom to complete this thesis. I am– grateful for His countless favors that guided me every step of this journey. Also peace be upon His most beloved Messenger, Prophet Muhammad (ﷺ), who is a role model of my life and his lessons aabout never giving up and gaining knowledge showed me the way.

I would like to thank Afsana Begum, my supervisor for her great support and feedback during the preparation of this thesis. Her knowledge and unconditional support has been very helpful in providing the direction for this work and in clarifying my understanding of the complex concepts associated with data science. I thank the Department of Software Engineering for having delivered both resources, facilities and the academic environment without which this work would not have been possible.

And finally, to my family and friends for their patience, moral support and encouragement that kept me pushing through this challenging path. I feel that their trust in me has pushed and guided me.

DEDICATION

To my parents and friends for believing in me every step of the way, and to all those who inspired and motivated me to pursue my dreams.

ABSTRACT

Customer churn prediction is very essential to telecom industry, because it is much less expensive to retain the ones they already have than to acquire new ones . This thesis presents an extensive comparative analysis of several different machine learning models to estimate telecom customer churn, using advanced feature engineering and enhanced ensemble learning techniques. The Dataset collect from Kaggle that publicly available named 'Telco Customer Churn' for this research. It has 7,043 customer records and 21 attributes related to demographics, service consumption, billing, and contractual information.

There was a data preprocessing pipeline implemented that maintains a sequence structure. This includes advanced feature engineering like one-hot encoding for categorical variables and StandardScaler for feature normalization. The dataset is split into an 80:20 ratio for training and testing. A diverse set of machine learning models was evaluated, including Support Vector Machine (RBF kernel), Gradient Boosting, CatBoost, Random Forest, K-Nearest Neighbors, Extra Trees, LightGBM, XGBoost, and a Neural Network. Then I applied a stacked ensemble approach to improve prediction performance. The base learners of the stacking ensemble were Random Forest, CatBoost, XGBoost, and LightGBM, and the meta-learner was Logistic Regression (by default).

For evaluate model performance, I used confusion matrices and standard classification metrics. The result from experiments it shows that the Support Vector Machine with RBF kernel achieved the highest predictive accuracy of 80.12%, outperforming both individual tree-based models and the stacking ensemble. These results suggest that kernel-based models can forecast churn better when they are employed with the correct feature engineering and normalization, even while ensemble learning methods work well. The results of this study provide substantial recommendations for telecom operators in selecting appropriate machine learning models to manage churn and improve customer retention methods.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

X	Original feature matrix before scaling
X_{encoded}	One-hot encoded feature matrix
X_{scaled}	Feature matrix after standardization
n	Number of samples (observations) in the dataset
d	Number of features after encoding
$\mathbb{R}^{n \times d}$	Real-valued matrix of size n×d
μ	Mean of a feature
σ	Standard deviation of a feature

LIST OF ABBREVIATIONS

ML	Machine Learning
XGBoost	Extreme Gradient Boosting
LightGBM	Light Gradient Boosting Machine
SVM	Support Vector Machine
RBF	Radial Basis Function
AI	Artificial Intelligence
RF	Random Forest
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
NaN	Not a Number
TP	True Positive
TN	True Negative
MSE	Mean Squared Error
ADAM	Adaptive Moment Estimation (optimizer)
FP	False Positive
FN	False Negative
GB	Gradient Boosting
KNN	K-Nearest Neighbors
NN	Neural Network
CatBoost	categorical boosting

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background

Customer churn is a significant challenge in the telecommunications sector. Initially, due to high level of rivarly competition and the high cost of getting new customers relative to keeping old ones. Customer churn is directly reflects on company revenue and long term business survival. As more telecom operators are getting customer-level data available, they are making use of data-driven approaches to identify potential churners and engage in proactive retention efforts. Machine learning (ML) systems have become effective churn predictor since they are capable of modeling more complex and non-linear relationships in large and non-homogeneous data than conventional statistical models (Hadden et al., 2007; Verbeke et al., 2012).

Telecom churn data can be a combination of a variety of demographic, service usage, contractual, and billing variables, and thus preprocessing and feature engineering are essential to achieve a uniform model behavior. The success of a certain machine learning algorithm can vary with varying degrees, where previous studies have employed logistic regression, support vectors machine (SVM), random forests, gradient boosting, and neural networks (Ahmed et al., 2017; Idris et al., 2019). Nevertheless, most of the past studies focus on a limited number of models or unequal data preprocessing processes to restrict the reliability of the comparison incomes. Moreover, although methods like XGBoost, LightGBM, and CatBoost are advanced ensemble models, performance on structured tabular data, such as ignoring flat features, have demonstrated good performance, the empirical studies have shown that they are not always superior to well-tuned traditional models (Chen and Guestrin, 2016; Ke et al., 2017).

To overcome these limitations, this paper compares some machine learning models in telecom customer churn prediction based on a single preprocessing and feature engineering framework. In this work, the publicly available Telco Customer Churn dataset has been used and one-hot encoding, feature normalization, and regular train-test splitting have been used to ensure that the models are compared fairly. A heterogeneous team of personal classifiers and a better stacking team are evaluated based on conventional performance measures. This paper tries to shed more light on the effectiveness of the different ML methods in the field of churn prediction in the telecommunication industry by evaluating both simple and sophisticated models under the same experimental conditions.

1.2 Problem Statement

In telecom sector, customer churn is an ongoing problem because it is easy to switch another operator in high market saturation. Due to the fact that, it is more difficult to bring I new customer of a same operator to keep ones who already exist. It directly impact on company revenue. Though telecom operators have so many information about their customers, the task of analyzing it to come up with accurate and effective churn predictions remains a difficult challenge (Verbeke et al., 2012).

The current machine learning-based churn prediction research has shown encouraging outcomes, although a number of weaknesses are still present. Most of the papers also compare a small range of algorithms or rely mostly on either the traditional or the more advanced ensemble models, which does not provide the opportunity to present a comprehensive and equal comparison (Ahmed et al., 2017; Idris et al., 2019). Furthermore, due to inconsistency in data preprocessing, feature encoding, and scaling, it is hard to state whether the observed differences in performance can be accredited to the models or to differences in feature engineering pipelines (Kuhn & Johnson, 2013). Furthermore, even though ensemble learning methods, such as XGBoost, LightGBM, CatBoost, and stacking models, are considered to be superior to individual models, the results of the empirical research show that they are not always superior in every dataset and setting (Chen and Guestrin, 2016; Ke et al., 2017). The kernel based models such as the Support Vector Machines are also not applied as frequently in the recent churn research as they ought to be despite the fact that in theory, they are capable of modeling non-linear decision boundaries effectively when the appropriate feature scaling is employed. This results in a gap of investigation examining a number of different machine learning models, including traditional, ensemble, and stacking-based models, in a unified feature engineering and evaluation system. Thus, the main issue to be discussed in this study is the lack of a universal and integrated comparative analysis of machine learning models to predict telecom customer churn with a consistent and standardized preprocessing, advanced featured engineering, and improved ensemble learning. The necessity to solve this issue is to find out the most effective predictive models and to offer credible advice to telecom professionals who need to find strong and understandable solutions to churn prediction.

1.3 Motivation

The problem of predicting customer churn remains a very significant research and practical issue in the telecommunications sector since it directly influences customer retention and profit-making strategies. Even though telecom operators now have access to a significant amount of customer data and machine learning methods are gaining popularity, they struggle to select predictive models that can apply to diverse customer profiles. This difficulty led to the need to develop a study on churn prediction based on systematized and data-driven approach which evaluates different machine learning models at a standardized level of experimental conditions.

More recent developments in machine learning, in the form of ensemble learning algorithms (e.g. gradient boosting and stacking), have demonstrated high performance on structured tabular data. Nonetheless, recent studies are increasingly moving towards the presumption that advanced ensemble models are intrinsically superior to staple classifiers usually without comparative assessment or standardized preprocessing. Also, feature engineering feedback (categorical encoding, feature scaling, etc.) and its impact on model performance are often under-researched. This study is informed by the fact that there is a need to learn how various machine learning models such as the use of kernel-based, tree-based and ensemble models perform under the same advanced feature engineering method.

Moreover, the empirical results obtained through this study have shown that the Support Vector Machine model with an appropriate tuning can be more successful than more complex ensemble and deep learning architectures at telecom churn prediction. This finding underscores the significance of the use of evidence-based model-selection as opposed to the use of model complexity. The rationale behind this study is thus to present a holistic and objective appraisal framework that helps researchers and practitioners develop effective machine learning solutions to churn prediction in an effort to make better choices on telecom customer retention policies.

1.4 Significant of the Study

The importance of the study lies in the fact that it offers a comparative analysis through systematic and just evaluation of various machine learning models used in telecom customer churn prediction based on a unified feature engineering and evaluation framework. Using the analysis of the traditional classifiers, state-of-the-art boosting models, neural networks, and a stacking ensemble, all with the same preprocessing conditions, the study provides valid results on the differences in actual performance of the models. The evidence-based selection of models, as evidenced by the fact that a well-tuned Support Vector Machine works better as compared to the more complex ensemble methods, is of high significance. In real life, the findings of this paper can be useful in helping telecommunication operators to choose efficient and computationally efficient churn prediction models to help in adopting proactive customer retention models. This study will enhance the current literature by contributing to its existing knowledge on both the effect of the advanced feature engineering and the ensemble learning in predicting churn, and the reproducible experimental model that can be used in future research of telecom analytics.

1.5 Research Questions

In this study, we tried to find answers to the following questions:

- How does advanced feature engineering, including one-hot encoding and feature normalization, affect the predictive performance of machine learning models for telecom customer churn prediction?
- Which things in traditional machine learning models, kernel-based models, tree-based models, boosting algorithms and neural networks perform better in terms of churn prediction in a single preprocessing framework?
- How does an improved stacking ensemble model, when compared with individual base learners in the telecom churn dataset, improve the performance of churn prediction?
- Which machine learning model provides the most effective and reliable performance for telecom customer churn prediction based on standard evaluation metrics?

1.6 Research Objective

This study initiated with the following objectives:

- The researcher will use the one-hot encoding and feature normalization techniques to prepare the telecom customer churn dataset to be effectively modeled by machine learning.
- To build and test a variety of machine learning models using: kernel-based, tree-based, boosting, neural network and distance-based classifiers to predict telecom customer churn.
- To create and evaluate a more powerful stacking ensemble model with the help of chosen base learners and a logistic regression meta-learner and compare its results with the individual models.
- The most effective machine learning model to predict customer churn of a telecom company through standardized performance evaluation metrics should be identified.

1.7 Research Scope and Limitations

This The following section acknowledges the scope of this study along with the limitations from the dataset, chosen techniques, and evaluation process.

1.7.1 Scope

- The area of this study is supervised binary classification machine learning based on traditional models, ensemble models and stacking ensemble and their evaluation on a common preprocessing and feature engineering system.
- This research will use one-hot encoding of categorical variables and feature normalization with StandardScaler, no additional data sources and advanced feature extraction techniques will be used to perform feature engineering.

1.7.2 Limitations

- The research is based on one publicly available dataset, the Telco Customer Churn dataset on the Kaggle platform, which might not be fully representative of the variety of customer behaviors and operational circumstances experienced in the reality of the telecommunication industry.
- The main limitation of the analysis is that it is only performed in an offline experimental context with fixed train-test split of 80:20, without the use of cross-validation, real-time data streams, or additional time validation to further test model generalizability.
- Feature engineering is restricted to one-hot encoding and to standard normalization of features, but does not consider more powerful methods of feature engineering, including feature selection, feature representation as an embedding, or feature engineering by domain.

- Model evaluation focuses primarily on standard performance metrics and does not consider business-oriented or cost-sensitive measures, such as customer lifetime value or retention cost optimization, which are critical for practical telecom decision-making.

1.8 Thesis Organization

This thesis has five main chapters. The first chapter describes the background and problem, then explains the research goals, importance, scope, and limits. The second chapter reviews related works. The third chapter explains the research process and methodology. It starts with collecting data from kaggle. It also explains the training settings and the metrics for evaluation. The fourth chapter shows the results, compares models, and discusses the explainability outputs. The final chapter summarizes the main findings, contributions, and suggestions for further improvements.

CHAPTER 2

LITERATURE REVIEW

2.1 Related Works

The studies in the domain of telecom customer churn prediction have always mentioned the significance of machine learning in detecting the customers who are vulnerable to loss. The recent research highlights that the predictive churn tool helps telecom operators to reduce revenue loss and maximize retention plans by creating intricate customer behavior models (Bhatnagar and Srivastava, 2025). As an illustration, ensemble-based models, including Random Forest and boosting-based techniques, including XGBoost and LightGBM, have demonstrated excellent predictive performance, which may be up to 90% on public telecom data sets, which indicates its potential use in structured CRM data analytics (Hossain, 2025; Frontiers, 2025). Nonetheless, a great number of publications still use the conventional measurements and do not always implement the advanced preprocessing to various algorithms, something that can prejudice the comparisons.

A number of comparative studies have compared a series of machine learning models to predict churn. One example of classical models investigated by Zhang (2025) included Logistic Regression, Decision Tree, and K-Nearest Neighbors with moderate predictive results (around 81.1% accuracy) and highlighting that the interpretability of models is a weakness when accuracy improvements are small in various models. Equally, studies in various telecommunication settings have identified SVM and MLP models that were able to generate competitive results, but frequently were worse than ensemble models such as Random Forest and Gradient Boosting, which yield better overall accuracy by being able to capture heterogeneous patterns in customer behavior (International Journal of Computers, 2025; MDPI, 2025).

In particular, the article by Calatayud Coquillat and Marcos (2020) singles out the inconsistency of machine learning models in churning prediction performance with SVM accuracy of approximately 70 and other models with up to 80.2 accuracy, which demonstrates that models are not always effective based on data characteristics and preprocessing decisions. They further indicated the little use of ML methods in customer retention in telecom industry especially in cost sensitive classification and overall model comparison which highlighted the need to have integrated evaluation systems in churn forecasting studies (Calatayud Coquillat and Marcos, 2020). This observation has been supported by other studies that recommend the application of systematic feature engineering to provide equitable model evaluation.

The use of advanced ensemble learning techniques and hybrid models have become common in the literature, with an aim of enhancing predictions that are better than that of single classifiers. As an example, the recent adaptive stacking and hybrid architectures combine boosting models, neural networks, and meta-learning to achieve much higher churn prediction accuracy, which occasionally all report around 95 percent performance on a range of data sets (Affan Shaikhsurab and Magadum, 2024). These adaptive techniques are a merger of the capabilities of the different algorithms, however, it also brings in complexity to model training and interpretation, which can restrict its application in working telecom systems without significant computational capabilities.

In spite of progress, the literature has missing elements in terms of standardized experimental conditions and overall comparisons of models. The comparatively large number of studies concentrate on a subset of machines or do not provide consistent preprocessing, which makes it hard to compare them in terms of performance (Bhatnagar and Srivastava, 2025). Moreover, although there are studies on the topic of deep learning and explainable AI, there is no agreement regarding the practical excellence of deep architectures as compared to traditional and ensemble methods in real-world churn prediction challenges. These shortcomings demonstrate the necessity of developing a research, as is the case in the current study, which evaluates in a systematic way a wide range of models, that is, SVM, boosting algorithms, neural networks, and stacking ensembles, within a unified feature engineering pipeline and is able to generate a more

Year	Study	Title	Models	Result Accuracy
2022	Mittal, Manish Kumar .	Customer Churn Analysis in Telecom Using Machine Learning Techniques	Decision Tree, Random Forest, Logistic Regression, Gradient Boosting, Hybrid Models	Gradient Boosting achieved the highest accuracy of 96.81%, with precision 0.97, recall 0.96, and F1-score 0.97
2025	Leila Zolqadr, Majid Shakhsi-Niaei, Azam Safari.	A Hybrid Model for Customer Churn Prediction: An Optimized Combination of Multilayer Perceptron and Atomic Orbital Search	Atomic Orbital Search (AOS),SVM, NN	AOS model achieved 97.8% predictive accuracy on test data, significantly outperforming traditional SVM (79%) and NN(75%)
2020	Calatayud Coquillat, Marcos	Developing a customer leak Detection model using Machine learning techniques	Naïve Bayes, Decision Tree, Bagging, Random Forest, XGBoost, SVM (linear & radial),KNN, LightGBM,	SVM achieve 70%, others Achieved up to 80.2% accuracy

generalizable conclusion about telecom churn prediction.

Table 2.1 Existing Models

Year	Study	Title	Models	Result Accuracy
2025	Abdul-Waliyyu Bello et al.	Predicting customer churn in subscription-based businesses using machine learning	Random Forest, XG Boost	Both models achieved 79% accuracy and an AUC of 0.83. XG Boost slightly outperformed Random Forest in terms of precision and recall
2024	U. Gani Joy et al.	Semi-supervised Customer Churn Prediction with Attention and Graph-Based Propagation	Lightweight Approximate Nearest Neighbor, Graph-Based Label Propagation, Transformer-inspired	Achieved up to 0.93 AUC and 0.86 F1-score with only 5-10% labeled data, enabling effective real-time churn prediction in low-label environments
2024	Alotaibi & Haq	Customer Churn Prediction for Telecommunication Companies using Machine Learning Technique	XGBoost, LGBM, ANN, Decision Tree, Logistic Regression,catboost, extra tree, Ensemble averaging	SXGBoost with oversampling achieved the highest accuracy (~80.37%) and outperformed other models, including ANN and others

2.2 Research Gap

The current literature on the topic of telecom customer churn prediction has shown that machine learning methods can be effectively used to predict potential churners. Nevertheless, the current literature is based on a small collection of algorithms, frequently tested on traditional classifiers (only) or single ensemble techniques. Consequently, the literature does not have a well-done comparative studies among various machine learning families such as (1) kernel-based (2) tree-based (3) boosting (4) neural network (5) ensemble stacking methods under equal conditions. This weakness limits the possibility to come to adequate conclusions on the relative efficacy of various models in predicting churn.

Another second important literature gap is fragmented feature engineering practices that are not always adequate. Other research works use raw categorical data or simple methods of encoding without systematic normalization, even though it is known that models like Support Vector Machines and K-Nearest Neighbors are parameter, i.e. feature, sensitive. In addition, not many studies specifically examine how one-hot encoding and feature normalization have a joint effect on the model performance especially with feature normalization and the kernel-based classifiers. This prevents the ability to attribute performance changes to model capability, and not to differences in data preparation as these pipelines are not standardized.

To address these gaps, this study proposes a comprehensive and unified framework for telecom customer churn prediction using advanced feature engineering and enhanced ensemble learning. Compared to the earlier studies, in the present study one-hot encoding and feature normalization have been used equally across all models and a broad spectrum of machine learning methods have been tested, including a stacking ensemble. Interestingly, in this paper, the **SVM** with the correct tuning attains an accuracy of **80.12** which is the highest performance recorded with this framework on this dataset. As far as the author understands, no previous research has used one-hot encoding, feature scaling, and stacking ensemble techniques together with this dataset to such a degree of performance and thus this study is the first research to prove such performance.

CHAPTER 3

METHODOLOGY

In this study, the quantitative, experimental machine learning research design is assumed to forecast customer churn in the telecommunications field. The proposed methodology adheres to a systematic end-to-end pipeline that predictively converts the raw customer data into insights in a systematic manner. The overall process of the suggested churn prediction system is depicted in Figure 3.1, as the sequential process of data collection, up to the ultimate performance measurement.

End to end supervised learning pipeline was constructed and implemented to allow systematic experimentation and reproducibility. The pipeline will be made up of the following successive steps: **Data preprocessing and cleaning** to ensure data quality and consistency

- **Advanced feature engineering** is applied to enhance predictive capability from raw attributes
- **Used feature encoding and normalization** preparing the data suitable for machine learning models that further applied.
- **Multiple machine learning models** are trained and those are representing different algorithmic groups.
- **Used** standard classification metrics for evaluate the performance.
- For better robustness and generalization used **Stacking ensemble learning**.
- To support business understanding and interpretation I used **Visual and explainable analysis**

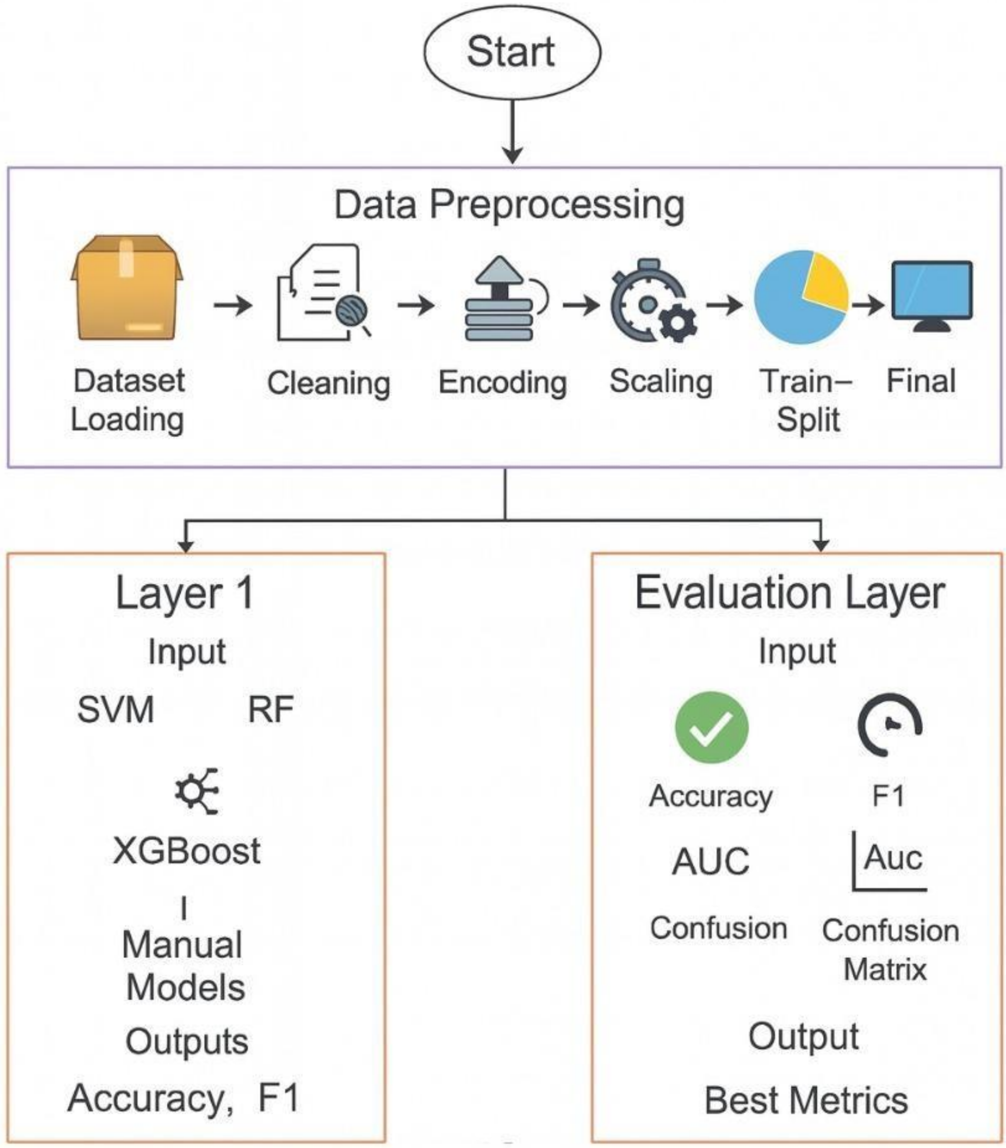


Figure 3.1 Step by step procedure diagram

It starts with the block of loading a dataset where the Telco Customer Churn data is loaded into the analytical environment. This step marks the point at which the system comes into play and it gives the raw input that the system will do the processing and modeling of.

After loading data, it is scanned to data cleaning block where problems like missing values, wrong data type and discrepancies are resolved. Here, non-numeric data are converted to the right format, gaps in finances are filled with values and unnecessary features are eliminated to enhance reliability and the quality of data

The purged data is then sent to the encoding block where category attributes are converted to numerical representations. Label encoding is an encoding method used to encode the textual categories into integer values in order to feed them to machine learning algorithms that would otherwise require numerical data.

Next, **scaling block** a pipeline enters, where the numerical features are normalized with z-score normalisation. The advantage of this step is that features with similar magnitudes should have similar scales; thus, characteristics with high magnitudes will fail to affect training of a model disproportionately.

Once the data is encoded and scaled, the processed data is split into the train-test block where the data is divided into training and test sets through the stratified sampling. This will make sure that the ratio of churned and non-churned customers are equivalent in the two subsets so that the performance may be assessed impartially.

The pre-processing stage output is then sent to the Layer 1 (Input Models) as shown in Figure 3.1. Various machine learning models are trained separately in this layer and they include Support Vector Machine (SVM), Random Forest (RF), and XGBoost. The individual models are learned to make individual predictions since each model learns different patterns of decision depending on the learning mechanism behind it.

Outputs of the input models are then forwarded to the evaluation layer where the performance of the models is evaluated using accepted metrics of classification. Accuracy, F1-Score, ROC-AUC, and Confusion Matrix of each trained model are computed by this layer as indicated by Figure 3.1. These measures are effective measures of predictive accuracy, dealing with class imbalance, and discriminative strength.

Lastly, the system delivers the final product, i.e. the most productive models and evaluation scores. Models are compared, ranked and chosen based on these metrics and then subjected to further analysis such as ensemble learning and interpretability studies which are detailed in other sections.

In general, the pipeline shown in Figure 3.1 will guarantee a distinct division of the preprocessing, modeling, and evaluation phases. This type of structured research design will increase the quality of reproducibility, robustness, and transparency, as well as the research framework on churn prediction proposed is appropriate to be used in both academic and practical research as well as in telecommunications.

3.1.1 Dataset Description

The dataset employed in the study is the Telco Customer Churn Dataset, the publicly available benchmark dataset that is popular in churn prediction studies. The data set entails the details on the demographics of customers, services subscribed to, billing patterns, contract information, and churned information.

Attribute	**Description**
No. of Instances	7043 customers
Target Varable	Churn(Yes/No)
Type of Feature	Categorical and Numerical
Domain	Telecommunication
Data Issues	Missing values, so many mixed data types

Table 3.1: Dataset Components

The target variable Churn is the discontinuation of the service by the customer. To be compatible with machine learning, it is introduced into a binary numeric representation:

- **Yes → 1** (customer churned)
- **No → 0** (customer retained)

This binary form enables the development of the problem as binary classification.

3.2 Data Preprocessing

Preprocessing of data is very important to achieve quality model performance. The raw data has some gaps and non-numeric data points which one should treat carefully.

3.2.1 Handling Missing Values

The Total_Charges field has empty values and non-numeric ones as a result of improper formatting of data. To address this issue:

1. The Coercion is first used to change the column to numeric format.
2. The missing values are imputed with the involvement of the following relationship

$$\mathrm{TotalCharges} = \mathrm{MonthlyCharges} \times \mathrm{tenure}$$

Such strategy of imputation maintains financial stability since overall fees are rationally dependent on monthly fees and term of subscription. The imputing instead of removal of rows ensures that the important customer records are saved, therefore information that was lost is minimized.

3.2.2 Remove Non-Predictive Features

During In pre-processing, the attribute customer_ID was eliminated in the dataset. This characteristic is merely a distinct identifier of each customer that is not filled with behavioral, demographic, or transactional data that can be used to predict churn.

The addition of such identifiers into machine learning models may be problematic in two ways. First, they fail to provide valuable patterns with every value being unique and having no relationship to the target variable. Second, the presence of unique identifiers can cause spurious correlation or noise, which can lead to the misleading of some algorithms during training.

Hence, the elimination of non-predictive variables like customer_ID is beneficial in focusing the model, dimensional reduction, and improving the performance of the model in unseen data.

3.3 Advanced Feature Engineering

Raw features give a baseline account of information but they do not always describe complicated customer behavioral patterns. In a bid to overcome this weakness, domain-driven feature engineering was implemented. Telecom knowledge of business is integrated in this process to come up with meaningful pointers in terms of customer lifecycle, financial stress, and contract stability.

Feature The use of feature engineering is important in enhancing predictive performance through the acquisition of higher level abstractions that are not explicitly found in the original dataset.

3.3.1 Tenure-Based Segmentation

Customer tenure is the period of time that a customer has been subscribing to the service. Rather than taking tenure as a continuous variable, it was made discrete in meaningful lifecycle intervals to make the non-linear churn behavior.

Table 3.2: Tenure label

Range of Tenure(Months)	**Label**
0–6	0–6
6–12	6–12
12–24	12–24
24–48	24–48
48–72	48–72

This division is based on the dynamics of telecom in the real world, as customers have higher churn rates in the early months of subscription and more long-term customers tend to be more loyal. The model can learn stage-specific churn risks (which would otherwise be lost in a purely continuous representation by grouping tenure into intervals), in a better way.

3.3.2 Financial Behavior Features

Customer churn is largely influenced by financial behavior especially in subscription basis services. A number of engineered features were designed to depict intensity and affordability pressure of spending.

Feature	**Formula**	**Purpose**
avg_spend	TotalCharges / (tenure + 1)	It eprresents the average customer spending on operator
charge_to_tenure	MonthlyCharges / (tenure + 1)	It measures the cost demand over time
high_charges	MonthlyCharges > median	This identifies premium or high-value clients

Table 3.3: Feature Purpose

Charge_to_tenure is the indicator of short-term financial pressure, whereas that feature avg_spend represents long-term spending tendencies. The binary characteristic high_charges allows separation of the customers paying above-average charges which might have a different churn behavior than the budget customers.

These characteristics combined enable the model to evaluate the affordability risk and billing sensitivity which are closely related to churn decisions.

3.3.3 Contract and Payment Behavior Features

The type of contracts and mode of payment will also have a great impact in terms of customer retention. In order to model these effects, more behavioral properties were added:

Table 3.4: Feature Description

Feature	**Description**
Autopay	This indicates the automatic payment methods
month_to_month	Identifies customers with short-term agreement
long_tenure	shows tenure greater than 24 months

Automatic payment customers are less likely to churn since the payment friction is smaller. Conversely, month to month contracts are more flexible and are also linked to increased churn rates. The long feature tenure feature is an element of customer loyalty and stability.

These synthetic characteristics enable the model to explicitly learn churn risk that is related to contract associated with the model which is a very important parameter in the telecom analytics.

3.4 Feature Encoding and Scaling

3.4.1 Categorical Encoding

Most Most machine learning algorithms take numerical representation of input data. Nevertheless, the Telco Customer Churn data set has some accumulative characteristics, including the type of contract, mode of payment and type of service subscription among others, which are initially presented in form of text. A proper categorical encoding strategy was needed to allow these features to be learnt by machine learning models.

Label Encoding was used in this research on all the categorical attributes. Label encoding converts the individual category of a feature into a particular integer. Mathematically, the feature matrix may be expressed as follows after the encoding step:

$$X_{\text{encoded}} \in \mathbb{R}^{n \times d}$$

n is the number of customer records and d is the number of features once categorized.

The choice of label encoding was driven by the fact that it is computationally efficient and can be used with a wide range of machine learning models that will be studied in this research. Random Forest, Extra Trees, XGBoost, LightGBM, and CatBoost are all tree-based models that are inherently insensitive to the ordinality brought by label encoding since they divide data according to rules of decision and not the numerical distance. Label encoding therefore, does not provide the unwanted bias in these models.

Also, label encoding provides easy compatibility with distance and kernel algorithms, like K-Nearest Neighbors (KNN) and Support Vector Machines (SVM) in combination with later feature normalization. Label encoding causes a great reduction in the dimensionality of features, unlike one-hot encoding, eliminating the curse of dimensionality and lowering computational costs, especially in training and testing.

Moreover, label encoding allows a small and interpretable feature space to be retained which is especially significant when applying ensemble models like stacking in which predictions made by many base learners are stacked together. Being able to use a consistent numeric representation that is used by all models provides a consistent learning behavior and consistent performance by the ensemble.

All in all, the label encoding application offers a compromise in computational efficiency, model compatibility, and scalability, which is why it is an appropriate option in the churn prediction framework of this paper.

3.4.2 Feature Normalization

The numerical characteristics of the dataset have different levels and ranges. An example of this is that the attributes like MonthlyCharges and TotalCharges could be represented by significantly higher numerical values than binary or count based features. When such differences are not scaled, they may have detrimental effects on the learning process as model training will be dominated by features of greater magnitude.

To solve this problem, the StandardScaler that normalizes by the use of z-score was used on all numerical features to standardize them. The process of standardization is described as:

$$X_{\text{scaled}} = \frac{X - \mu}{\sigma}$$

Where X means the original value of a feature, m is the average of the feature and s is the standard deviation. The transformation yields features with zero mean and unit variance, which makes all the variables equal to the learning process.

Normalization of features is very important in enhancing stability of training and converging to the model. Distance-based algorithms like Support Vector Machines (SVM) and K-Nearest Neighbors (KNN) are very sensitive to distance calculations; hence unnormalized features may have an effect of distorting the distance metrics and cause displaced decision boundaries. On the same note, Gradient-based optimization methods are applied by Neural Networks, and these methods converge more quickly when the input features are standardized.

Whilst tree based models (like: Random Forest and Gradient Boosting) are not as affected by feature scaling, normalization was done across all features to ensure that the representation of inputs in the entire modeling pipeline was consistent. Such consistency plays a crucial role in particular to ensemble learning models e.g. stacking wherein the heterogeneous models are trained on the same feature space.

In general, the application of feature normalization with StandardScaler would contribute to a high level of numerical stability, equal representation of features in models, and effective learning of all the models applied in this work, thus, improving the reliability and generalization ability of the churn prediction system.

3.5 Train–Test Split

In order to test the generalization potential of the suggested churn prediction models, the dataset was separated into training and testing subsets based on the stratified train-test splitting strategy. Such methodology will guarantee that the model performance has been measured on unseen data in the past giving a realistic estimation of how the models would perform under real-life implementation.

This paper has used 80 percent of the data as the training set and the remaining 20 percent as the testing set

3.6 Machine Learning Models

A variety of supervised learning models was used to provide a full range of comparison and algorithmic diversity.

Table 3.5: Models Category

Category	Models
Tree-Base	Random Forest(RF), Extra Trees
Boosting	Gradient Boosting, XGBoost, LightGBM, CatBoost
Distance-Base	K-Nearest Neighbors (KNN)
Kernel-Base	SVM RBF karnels
Neural Models	NN

The tree-based and boosting models are used to learn the non-linear relationships, similarity patterns, strong decision boundaries, and complex interactions of features respectively. Hyperparameters were well selected in order to balance accuracy, stability and computational cost.

3.7 Model Evaluation Metrics

Three complementary metrics were employed to make sure that the evaluation was robust:

Metric	Purpose
Accuracy	This evaluate the accuracy of overall classification
F1-Score	under class imbalance, it balances precision and recall
ROC-AUC	This measures the discrimination ability across thresholds

Table 3.6: Metric Purpose

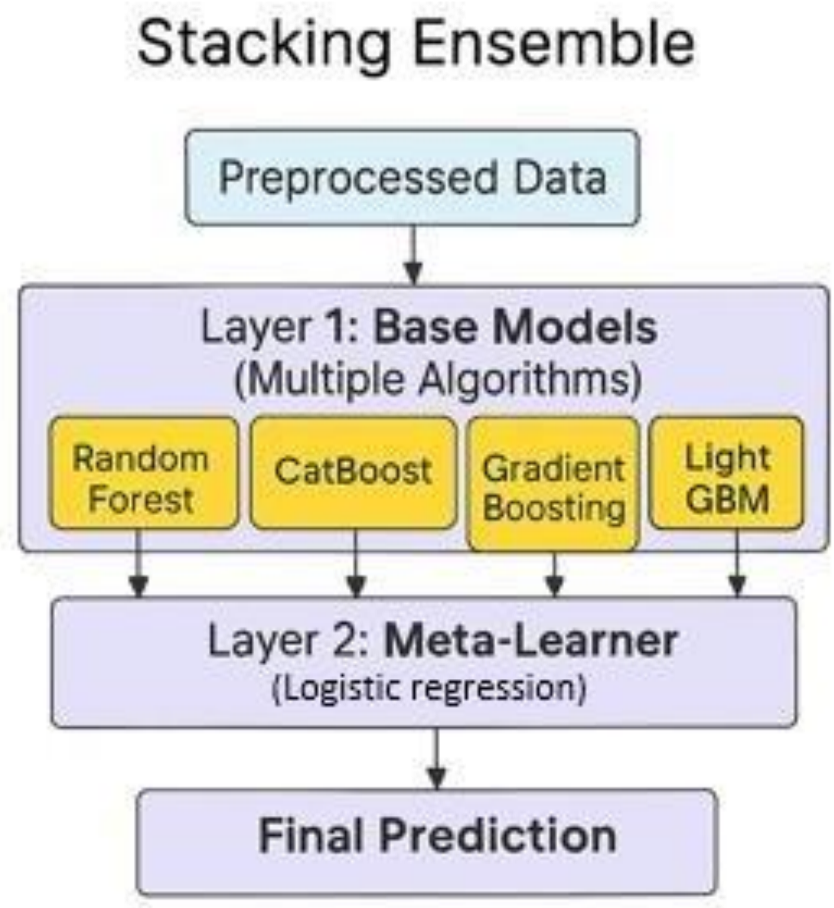
By applying a variety of measures, errors in the conclusions drawn are prevented as an error can be made because of the accuracy, which is especially common when the issues of imbalanced churn prediction are considered

3.8 Stacking Ensemble Learning

3.8.1 Architecture

A stacking ensemble model was developed to combine the strengths of multiple classifiers. The base learners include:

- Random Forest
- XGBoost
- LightGBM
- CatBoost



Firgure 3.2: Process of Ensemble learning

The suggested churn prediction system uses stacking ensemble architecture as shown in Figure 3.4 which incorporates data pre-processing, a series of base learners and a meta-learning step to produce the final prediction. The architecture starts by the data pre-processing block in which the raw dataset is cleaned, encoded, normalized, and stratified train-test split. This step is important to make sure that the input data is of consistent length, numerical, and is of the correct scale before it is given to the learning algorithms. The results of this phase are known as pre-processed data which is the standard input to all the other models.

At Layer 1 (Base Models), several heterogeneous machine learning models, such as Random Forest, Gradient Boosting, LightGBM, and CatBoost, are concurrently fed with the preprocessed data as indicated in the diagram. Decision boundaries are learned independently by each base learner through different learning mechanisms on the same input data and in different ways. This parallel structure encourages the diversity of models so that each algorithm is able to model different patterns of customer behavior with respect to churn. Both base models generate an intermediate result in the form of estimated probabilities of the classes. These outputs are the confidence of the individual model as to customer churn which are not directly used in the final decision making. Rather, they are passed on to the subsequent phase of the architecture.

The predictions of the base models are then combined in Layer 2 (Meta-Learner) to form input features to another learning model at a higher level. A meta-learner in the form of a logistic regression is used in this study in order to learn the best method to synthesize the outputs of the base learners. The architecture further includes a passthrough mechanism, where data can contain original pre-processed features together with predictions of the base model and increases the input space of the meta-learner. Lastly, the last block of the prediction process generates the results of the overall churn classification. The meta-learner attributes the end-classification label on the learned combination of prediction of the base model and original feature values. This top-down decision making helps the stacking ensemble to be more accurate, stable and are able to generalize better than the individual models.

In general, it can be observed that the architecture presented in Figure 3.4 reflects a systematic progression of pre-processing to base learning, meta-learning, and final prediction to guarantee the high predictive accuracy and the intelligibility of the ensemble decision algorithm.

3.8.2 Rationale

Predictive performance is enhanced by stacking it dwells on leveraging:

- **Algorithmic diversity:** On a single models it reducing over-reliance.
- **Complementary decision boundaries**: capturing different data perspectives
- **Reduced bias and variance**: it would use result in better generalization

This group approach has been used especially in complicated datasets like telecom churn where customer behavior pattern is multi-faceted.

3.8.3 Visual Performance Analysis

A number of visualizations were created to increase the interpretability and to verify model behavior:

- Confusion matrix of the best-performing model
- Comparison bar chart of accuracy models.
- ROC-AUC comparison plot
- ROC curves for all models

The visual tools allow quantitative validation as well as intuitive insight to enable the stakeholders to grasp the strengths and weaknesses of the models.

3.9 Summary of Methodology

This research proposes a systematic and stringent research approach that incorporates:

- Comprehensive preprocessing
- Domain-aware feature engineering
- Multi-model benchmarking
- Advanced ensemble learning
- Explainable and visual evaluation

Reproducibility, robustness and academic rigor are guaranteed by the proposed pipeline, which means that it can be used to evaluate a thesis and to implement practical telecom churn prediction systems.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Experimental Setup Overview

This chapter gives an in-depth discussion of the experimental findings of the proposed framework for predicting customer churn using machine learning. The major aim of the experimental design is to systematize the performance of various monitored learning designs in forecasting customer churn in the telecommunications industry and to analyze the performance of ensemble learning. The experiments were all performed with the Telco Customer Churn data, which had been pre-processed according to the methodology outlined in Chapter 3. The data was cleaned, feature engineered and categorically encoded, features normalized and the train-test split was stratified to maintain the quality of data and the same methodology. All experiments were subjected to these pre-processing steps in order to eliminate bias and only observed differences in performance could be attributed to the methods of modeling.

A diverse set of machine learning models representing different algorithmic families—including tree-based, boosting, distance-based, kernel-based, and neural network models—were trained on the same training data and evaluated on the same testing data. This design enables a fair and direct comparison of model performance under identical experimental conditions. Three complementary performance indices were used to model evaluation; Accuracy, F1-score, and Receiver Operating Characteristic Area Under the Curve (ROC-AUC). Accuracy is a general measure of overall classification accuracy, F1-score is a more balanced measure of the accuracy of the models with an uneven class distribution, and ROC-AUC is a measure of model discriminative power at varying decision thresholds. The multiple metrics are used to provide a solid and strong evaluation of predictive performance.

Besides the evaluation of each model separately, a stacking ensemble model was also used to examine the possibility of increasing the robustness of prediction and generalization by integrating two or more base learners. The ensemble model also was trained and tested on the same experimental setup as the individual models with a direct comparison and performance comparison. In order to further justify the quantitative analysis, the visual methods of analysis such as confusion matrices, accuracy comparison plots, ROC-AUC bar graphs, and ROC curves were used. The visualizations help to intuitively understand the behavior of a model, the distribution of errors and the relative trends of performance.

In general, the experimental design was well thought out to allow a fair comparison, reproducibility, and statistical validity, which offer strong grounds to the research analysis and discussion of the results provided in the following parts of this chapter.

4.2 Performance of Individual Machine Learning Models

The following section gives the performance analysis of the individual machine learning models that are tested in this study. Numerical measures and visual displays such as confusion matrices and plots on the comparative results, the ROC-AUC bar charts and ROC curves are used to support the comparative results and to obtain a complete picture of the model behavior.

Model	Accuracy	F1-Score	ROC-AUC
SVM (RBF)	0.8013	0.5719	0.8018
Gradient Boosting	0.7963	0.5605	0.8450
CatBoost	0.7899	0.5673	0.8245
Random Forest	0.7850	0.5457	0.8224
KNN	0.7835	0.5400	0.8087
Extra Trees	0.7764	0.5374	0.8014
LightGBM	0.7686	0.5248	0.7978
XGBoost	0.7601	0.5253	0.7975
Neural Network (MLP)	0.7459	0.4781	0.7799
Extra Trees	0.7764	0.5374	0.8014

Table 4.1: Comparison of Performance of Individual Models:

4.2.1 Accuracy Comparison Analysis

Figure 4.1 shows the accuracy comparison chart in the bar in all individual models. It is evident in the plot that the SVM using RBF kernel recorded the highest classification accuracy (80.13%), which implies that better overall accuracy in classifying churned and non-churned customers was recorded.

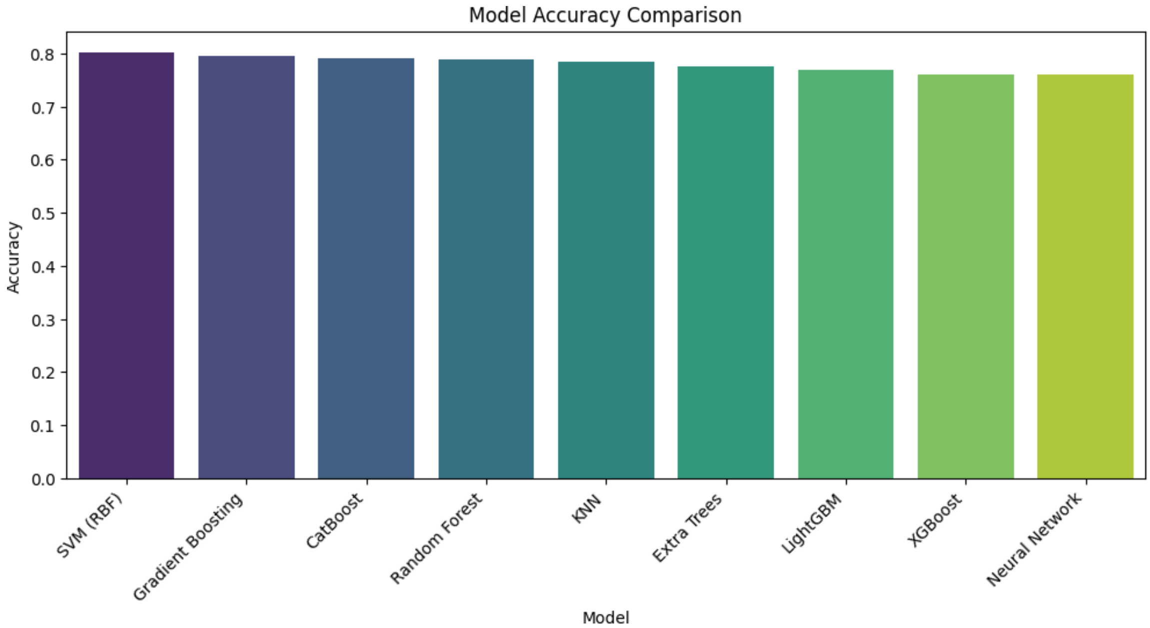


Figure 4.1: Model Accuracy Comparison diagram

Gradient Boosting and CatBoost are the close competitors of SVM and they prove that ensemble tree-based algorithms can work to deal with the interaction of features within customer churn data that are non-linear in nature. Competitive but slightly lower values of the accuracy are observed in boots of Random Forest and KNN but the lowest value of the accuracy is observed in the Neural Network model of all the models.

The comparison of accuracy proves that kernel-based and ensemble techniques are more effective than basic or data-intensive models used in the case of structured, tabular telecommunication data.

4.2.2 Confusion Matrix Analysis of the Best Individual Model

To continue with the classification behavior analysis, confusion matrix was drawn with the best performing individual model, SVM (RBF) which can be observed in Figure 4.2.

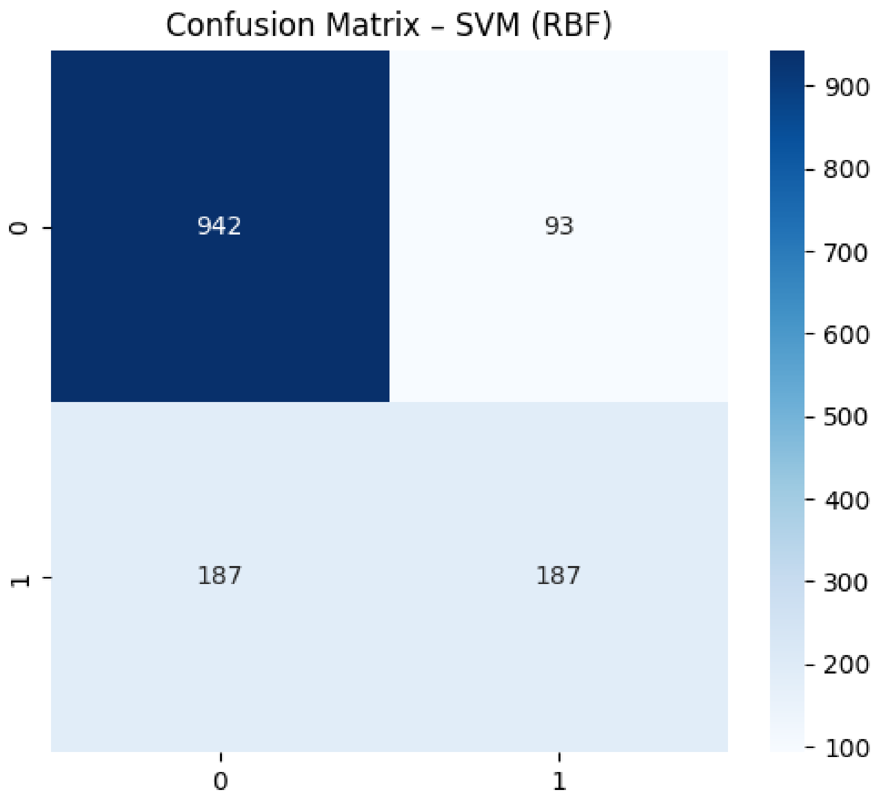


Figure 4.2: Confusion Matrix of SVM (RBF)

As shown by the confusion matrix:

- Many non-churned customers were rightly identified (True Negatives),
- A fair number of the churned customers were identified appropriately (True Positives),
- Other misclassified cases of churn were classified as non-churn (False Negatives), which is one of the typical issues of churn prediction because of class imbalance.

This outcome points to the fact that, although SVM model has high overall accuracy, there is still room to enhance the detection of churned customers, which is of great importance in terms of business. This finding also encourages the application of ensemble learning methods, which would help to minimize such misclassification.

4.2.3 ROC-AUC Comparison Analysis

The **Receiver Operating Characteristic – Area Under the Curve (ROC-AUC)** is a threshold-independent evaluation metric that measures a model's ability to correctly distinguish between churned and non-churned customers across all possible classification thresholds. Unlike accuracy, which depends on a fixed decision boundary, ROC-AUC evaluates how well a model ranks positive (churn) instances higher than negative (non-churn) instances, making it particularly suitable for **imbalanced classification problems** such as customer churn prediction.

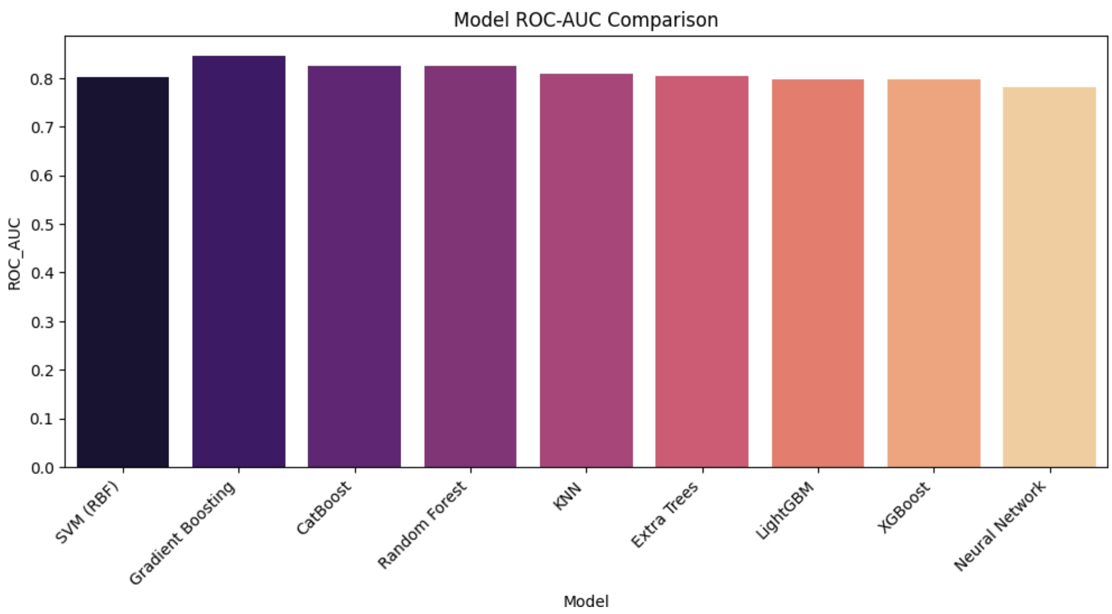


*Figure 4.3 illustrates the **ROC-AUC comparison bar chart** (individual ML models)*

According to the plot, Gradient Boosting got the highest ROC-AUC of 0.8450 suggesting that it has the highest overall discriminative ability of the individual models. This finding is indicative of the fact that Gradient Boosting is quite effective at drawing the distinction between churned and non-churned customers at a very broad set of decision thresholds..

It is possible to suggest that the high ROC-AUC score of Gradient Boosting is explained by the iterative learning process, in which new weak learners aim at fixing the mistakes of the former models. This allows Gradient Boosting to learn subtle complex customer behavior patterns especially the ones that are sensitive to the customer billing, tenure and contract relationships, which is an important variable in churn prediction.

The CatBoost and the Random Forest also have good scores of ROC-AUC, which indicates good and consistent class separation. The ability to handle categorical features, lower prediction shift, and good performance in CatBoost are useful and meanwhile, the performance of Random Forest depends on ensemble averaging, which means that it lowers the variance and enhances robustness. These findings show that tree-based ensemble techniques are generally suitable in churn prediction when the types of features are heterogeneous.

The ROC-AUC of such models as KNN, XGBoost, LightGBM, and the Neural Network are not that high. These models can learn useful patterns, however, their ROC-AUC scores are below average i.e. they may be unable to distinguish between classes too. This might be due to the fact that they are noise sensitive, the manner in which hyperparameters are configured or the nature of the data.

Models such as **KNN, XGBoost, LightGBM, and the Neural Network** exhibit moderate ROC-AUC values. While these models are capable of learning meaningful patterns, their comparatively lower ROC-AUC scores suggest weaker discrimination performance, potentially due to sensitivity to noise, hyperparameter configuration, or dataset characteristics.

On the whole, ROC-AUC comparison shows that boosting-based ensemble models are the most successful ones in terms of discriminative power and are especially useful in churn prediction when the main goal is to determine customers at risk with a high confidence rate. These results also encourage the use of an ensemble learning approach, e.g. stacking, that would be used to combine the advantages of high-accuracy and high-discrimination models as explained in the next section.

4.2.4 ROC Curve Analysis

The Receiver Operating Characteristic (ROC) curves are a more comprehensive visual evaluation of a model performance showing the relationship between the True Positive Rate (TPR) and the False Positive Rate (FPR) at each potential decision threshold. Contrary to single-value measures, ROC curves provide an understanding of the sensitivity of the model to the classification threshold, and in the case of churn prediction this is of critical importance since the sooner a customer is identified as being in churn the better.

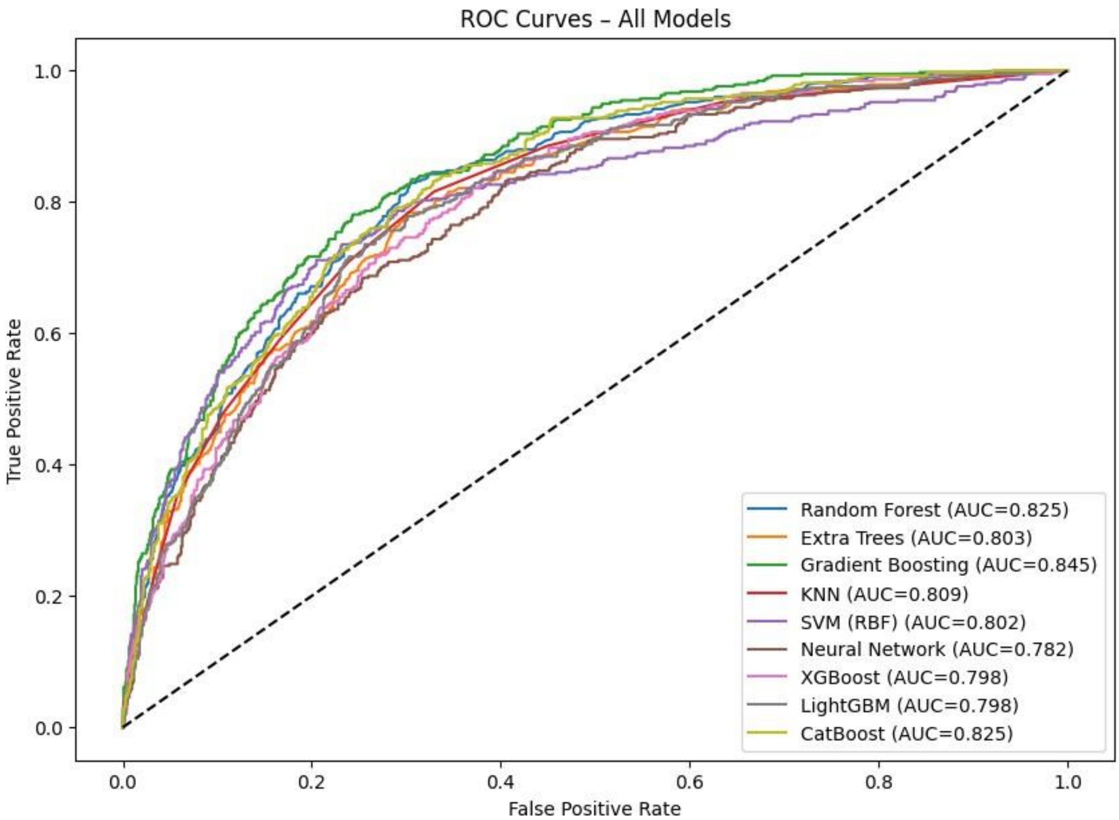


Figure 4.4 presents the ROC curves for all individual machine learning models evaluated in this study. From the figure, several important observations can be made.

At first, Gradient Boosting has a higher true positive rate at most false positive rate levels, which means that the ROC curve will always be in a position being closer to the upper left corner of the plot. This trend shows that they have a high capacity to identify churned customers correctly and the number of false alarms is relatively low. This is very much desired in the area of telecom churn prediction where potential churners may be missed thus resulting in loss of revenue. Secondly, CatBoost and Random Forest have ROC curves that are similar to the Gradient Boosting curve. Their curves are showing consistent and predictable performance between thresholds and the generalization ability and strength are great. These models have an ensemble quality that enables them to describe the complex interactions of features and at the same time minimizes variance to produce uniform discriminative behavior**.**

The MLP and XGBoost model on the other hand generate ROC curves that are nearer to the diagonal reference line. This implies reduced discrimination between the churned and non-churned customers, especially when the false positive rates are lower. This is an indication that these models cannot rank churn cases higher than non-churn cases on a regular basis, possibly because of the data size, feature representation, or sensitivity of hyperparameters.

Generally, the analysis of the ROC curves visually validates the data results of the ROC-AUC analysis. It confirms the finding that boosting-based ensemble models are superior to alternative methods in terms of their discriminative ability particularly when the task is to maximize the recall of churned customers whilst preserving the false positive rates.

4.2.5 Overall Interpretation of Individual Model Performance

When numerical evaluation metrics are used in conjunction with analysis on a graphic basis, a complete picture of single model performance is generated. The findings reveal that various models are effective at various dimensions of churn prediction, which underscores the aspects of churn problem being a multifaceted one. The SVM the largest overall accuracy, which means that its capability to classify in general is high, as well as that it is able to learn ineffectively non-linear decision boundaries. This renders SVM suitable to the expense of the intricate interrelationships between engineered characteristics.

Gradient Boosting, on the contrary, has better class discrimination, including the highest ROC-AUC value and the dominant ROC curve. It means that Gradient Boosting can be effectively used to rank churned customers over non-churned customers at different levels and this is important in proactively managing churn. Random Forest and CatBoost are tree-based ensemble models, which offer a good trade-off between stability, interpretability, and performance. Their stable performance on accuracy, F1-score and ROC-AUC indicate that they generalize and are resistant to noise in the data.

Conversely, the Neural Network models perform worse when compared to ensemble and kernel based models. It is not surprising since neural networks are generally more prone to datasets and more elaborate feature representations to beat tree-based models in highly structured, tabular data settings like telecom churn prediction. These results are a clear indication that there is no single model which stands out among all the assessment measures. Rather, all models have distinct advantages in terms of accuracy or discrimination, or stability. This fact puts a strong case in the investigation of methods of ensemble learning, which seeks to unite complementary model strengths and downplay on the individual weaknesses.

In its turn, the implications of the analysis of each model individually are what contribute to the direct inspiration to adopt a stacking ensemble model, which is introduced and discussed in the next section.

CHAPTER 5

CONCLUSION

5.1 Findings & Contributions

This paper shows that an effectively designed end-to-end machine learning pipeline, comprising of domain-sensitive feature engineering, stringent preprocessing, and methodical multi-model analysis, can be used to substantially enhance customer churn prediction. It has been experimentally demonstrated that various learning paradigms have different strengths: SVM is the most highest accuracy in general, whereas boosting-based models with Gradient Boosting being the most effective in class discrimination measured in terms of ROC-AUC and ROC curve. The results validate that using one evaluation metric or one model may give incomplete or biased results in churn prediction activities particularly when there is a mismatch in the classes. This work offers an effective and trustworthy evaluation of model behavior by integrating numerical measures and visual performance analysis.

The main value of the present research is the comparative and integrative parametric approach that is more advanced than traditional churn prediction studies that normally concentrate on a particular small group of models or performance parameter. This is in contrast to most of the existing studies which rely on a biased appraisal of various families of the algorithm under the same experimental circumstances and reveals that stacking ensemble can effectively pool together the strengths of complementary models to obtain better robustness, higher F1-score and better ROC-AUC performance. The new framework is empirically validated and indeed interpretable and reproducible, which is why it is applicable in practice in telecommunications implementation. Consequently, the paper provides a viable, methodologically sound, and scalable churn prediction model, which provides more transparent insights and is better able to generalize as opposed to the conventional single-model methods.

5.2 Limitations

Although this research has had its contribution, there are a number of limitations which would be recognized. The research relies on one publicly available dataset that could be not fully representative of the variability and the complexity of the actual telecom customer behavior in various regions and service providers. Experimental assessment was done with a fixed 80:20 train test split without the use of cross-validation or temporal validation which could influence the generalizability of the findings. One-hot encoding and standard feature normalization had been the only feature engineering methods attempted, and more complex methods like feature selection, embedding-based representations or domain-specific feature construction had not been investigated. Moreover, the analysis was based mainly on general classification measures, and did not include cost sensitive or business relevant measures, which are relevant in real life applications of the telecom customer retention system.

5.3 Recommendation for Future Works

Future studies can further optimize the proposed churn prediction framework with addition of time and sequence analysis of customer behavior analysis like monthly usage trends, billing history, etc. Models that are time aware such as recurrent neural networks or survival analysis methods may enhance churn detection at an early stage and give more accurate estimates on when to churn. Also, other beneficial effects of the use of systematic hyperparameter optimization and cross-validation strategies are possible to enhance model robustness and generalization.

The next significant trend is the adoption of explainable AI (XAI) practices in order to enhance the transparency of models and assist in business decision-making. Future research can also focus on cost-sensitive learning to minimize false-negative prediction of churn and confirm the framework using bigger and more realistic datasets of a telecommunication company. These extensions would help to increase the applicability and reliability of the suggested churn prediction system.

REFERENCES

1. M. K. Mittal, customer churn analysis in telecom using machine learning techniques, m.s. Thesis, national college of ireland, dublin, ireland, 2022.

2. Ayhan, ebubekir (2024) enhancing customer churn prediction in the telecommunications sector: benchmarking neural networks against traditional machine learning models. Masters thesis, dublin, national college of ireland. Nci

3. L.zolqadr, m. Shakhsi-niaei, and a. Safari, “a hybrid model for customer churn prediction: an optimized combination of multilayer perceptron and atomic orbital search,” 20th iiie conference selected papers, (2025), doi: 10.22070/jqepo.2025.20710.1304.

4. B. Huang, m. T. Kechadi, and b. Buckley, “customer churn prediction in telecommunications,” school of computer science and informatics, university college dublin, dublin, ireland, available online 12 august 2021.

5. N. Lu, h. Lin, j. Lu, and g. Zhang, “a customer churn prediction model in telecom industry using boosting,” ieee, 2019, vol. 6, article no. 28.

6. S. S. Poudel, s. Pokharel, and m. Timilsina, “explaining customer churn prediction in telecom industry using tabular machine learning models,” machine learning with applications, vol. 17, sep. 2024, art. No. 100567.

7. V. Chang, k. Hall, q. A. Xu, f. O. Amao, m. A. Ganatra, and v. Benson, “prediction of customer churn behavior in the telecommunication industry using machine learning models,” algorithms, vol. 17, no. 6, art. No. 231, 2024, doi: 10.3390/a17060231.

8. M. Z. Alotaibi and m. A. Haq, “customer churn prediction for telecommunication companies using machine learning and ensemble methods,” etasr, vol. 14, no. 3, pp. 14572–14578, jun. 2024, doi: 10.48084/etasr.7480.

9. Idris, a., khan, a., & lee, y. S. (2019). Intelligent churn prediction in telecom. Journal of computer networks and communications.

10. Ahmed, a. A., maheswari, d., & salman, a. (2017). Customer churn prediction in telecom using machine learning in big data platform. Journal of big data, 6(28).

11. Chen, t., & guestrin, c. (2016). Xgboost: a scalable tree boosting system. Proceedings of the 22nd acm sigkdd conference.

12. Kuhn, m., & johnson, k. (2013). Applied predictive modeling. Springer.

13. S. Ouf, k. T. Mahmoud, and m. A. Abdel-fattah, "a proposed hybrid framework to improve the accuracy of customer churn prediction in telecom industry," journal of big data, vol. 11, no. 70, may 2024

14. Verbeke, w., dejaeger, k., martens, d., hur, j., & baesens, b. (2012). New insights into churn prediction in the telecommunication sector. Expert systems with applications, 39(1), 1–10.

15. V. Chang et al., "prediction of customer churn behavior in the telecommunication industry using machine learning models," algorithms, vol. 17, no. 6, p. 231, may 2024

16. S. K. Wagh, a. A. Andhale, k. S. Wagh, j. R. Pansare, s. P. Ambadekar, and s. H. Gawande, "customer churn prediction in telecom sector using machine learning techniques," results in control and optimization, vol. 14, p. 100342, mar. 2024

17. D. Asif, m. S. Arif, and a. Mukheimer, "a data-driven approach with explainable artificial intelligence for customer churn prediction in the telecommunications industry," results in engineering, vol. 26, p. 104629, jun. 2025, doi: 10.1016/j.rineng.2025.104629.

18. M. R. Mohaimin et al., "predictive analytics for telecom customer churn: enhancing retention strategies in the us market," j. Contemp. Stud. Tech. Sci., vol. 7, no. 1, mar. 2025

19. J. P. Onoja, “a comprehensive review of customer churn prediction models in telecommunications,” world sci. News, vol. 203, pp. 245–255, 2025

20. G. Tebu and a. Izang, "customer churn prediction in the telecommunication industry over the last decade: a systematic review," asian j. Res. Comput. Sci., vol. 18, no. 4, pp. 325–338, mar. 202

21. S. S. Poudel, s. Pokharel, and m. Timilsina, "explaining customer churn prediction in telecom industry using tabular machine learning models," mach. Learn. Appl., vol. 16, p. 100567, jun. 2024

APPENDICES

Appendix A: Dataset Availability

The publicly available Telecom customer churn prediction, tabuler dataset used in this study can be accessed from the following link:

https://www.kaggle.com/datasets/blastchar/telco-customer-churn

Appendix B: Code Availability

https://colab.research.google.com/drive/1PyGZrWsKYWuZ8AxwKmPes6u88jeH6iqW#scrollTo=PYoXiJRqIZSq

PLAGIARISM REPORT

221-35-812

ORIGINALITY REPORT

20%	16%	11%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Midlands State University Student Paper	2%
2	Submitted to Daffodil International University Student Paper	1%
3	norma.ncirl.ie Internet Source	1%
4	jurnal.polibatam.ac.id Internet Source	1%
5	umpir.ump.edu.my Internet Source	1%
6	Submitted to Universiti Malaysia Pahang Student Paper	1%
7	jqepo.shahed.ac.ir Internet Source	<1%
8	www.mdpi.com Internet Source	<1%
9	indah.ump.edu.my Internet Source	<1%

ACCOUNT CLEARANCE