Customer Churn Prediction In Telco: A Machine Learning Approach

Bachelor Thesis

Author: Yusuf Shehadeh (Student ID: 7395116)
Supervisor: Prof. Dr. Markus Weinmann
Co-Supervisor: Sercan Demir

Faculty of Management, Economics and Social Sciences
University of Cologne

July 29, 2024

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Schriften entnommen wurden, sind als solche kenntlich gemacht. Die Arbeit ist in gleicher oder ähnlicher Form oder auszugsweise im Rahmen einer anderen Prüfung noch nicht vorgelegt worden. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Die Strafbarkeit einer falschen eidesstattlichen Versicherung ist mir bekannt, namentlich die Strafandrohung gemäß § 156 StGB bis zu drei Jahren Freiheitsstrafe oder Geldstrafe bei vorsätzlicher Begehung der Tat bzw. gemäß § 161 Abs. 1 StGB bis zu einem Jahr Freiheitsstrafe oder Geldstrafe bei fahrlässiger Begehung.

Yusuf Shehadeh

Köln, den 29.07.2024

Abstract

This thesis presents an innovative approach to predicting customer churn in the telecommunications industry by integrating advanced machine learning techniques. The study addresses the critical need for accurate churn prediction, enabling telecom companies to devise targeted retention strategies and maintain a competitive edge. A comprehensive dataset comprising various customer attributes was meticulously prepared and analysed to uncover significant predictors of churn.

Key machine learning models, including RandomForest, XGBoost, LightGBM and Decision Trees were employed and subjected to hyperparameter tuning to enhance their performance. The stacking model, combining these base models, demonstrated superior predictive recall, precision and F1 score, significantly outperforming the baseline model. Evaluation metrics such as recall, precision, F1 score, ROC AUC and accuracy were used to rigorously assess model performance. The findings revealed features such as age, payment delay, usage frequency and the number of support calls are strong predictors of churn. These results underscore the importance of personalised retention strategies, service improvements, and product customisation to mitigate churn. The study also highlights the potential of advanced machine learning models in improving churn prediction.

Future research directions include incorporating diverse datasets to enhance model generalisability, exploring enhanced feature engineering techniques and testing more sophisticated machine learning models. The study advocates for continuous model updating and the use of automated hyperparameter tuning to adapt to evolving customer behaviours and market conditions.

The proposed approach contributes significantly to the existing body of knowledge in churn prediction and provides practical insights for telecom companies to enhance customer retention strategies, thereby fostering long-term business sustainability and growth.

Contents

List of Figures

List of Tables

1 Introduction

The telecommunications industry, characterised by intense competition and rapid technological advancements, faces the continual challenge of customer retention. In this context, predicting customer churn - subscribers discontinuing their service - is critically important for maintaining business stability and growth. Effective churn prediction enables telecom companies to devise targeted interventions to retain customers, which is vital for sustaining revenue and a competitive advantage.

Existing research in churn prediction has extensively utilised machine learning techniques to analyse customer behaviour and identify potential churners. Innovations have spanned from traditional statistical methods to complex machine learning models like deep learning. Studies have highlighted the effectiveness of integrating artificial intelligence with customer relationship management systems, employing big data analytics, and exploring emerging technologies such as IoT and blockchain to enhance predictive capabilities.

While these studies have significantly advanced churn prediction, they often focus on singular or isolated approaches, potentially overlooking the synergistic potential of combining different methodologies. Moreover, there remains a need for improved feature engineering and model evaluation techniques to enhance the accuracy and applicability of predictive models in diverse operational settings.

This leads to the pivotal research question:

How can integrating and enhancing various machine learning techniques improve the recall and robustness of churn prediction models in the telecommunications industry?

This study aims to address these gaps by proposing a novel, integrated approach to churn prediction. By combining multiple advanced machine learning techniques, the study seeks to develop a more robust and accurate predictive model that is adaptable to varying market conditions and customer behaviours.

The methodology involves developing a hybrid model that integrates Gradient Boosting Machines (GBMs), Random Forests and other models. This approach leverages the strengths of each model to handle different aspects of the data and prediction process. Advanced hyperparameter tuning techniques are implemented to refine model performance, enhancing the model's ability to learn relevant patterns effectively.

Model stacking is utilised to further improve predictive recall by combining the outputs of multiple models, thereby harnessing their collective predictive power. The study employs a comprehensive set of performance metrics to rigorously evaluate the model's effectiveness. Additionally, advanced cross-validation techniques are used to ensure the model's robustness and generalisability across different datasets.

2 Related Work

In this section, we explore a curated selection of scholarly articles and research studies that have contributed to the field of churn prediction within the telecommunications industry.

2.1 Performance Analysis of QoS Parameters and Churn Prediction Model Development for MNOS in Nigeria

The study by Summaila et al.(Summaila & Stephen, 2023) represents a significant contribution. Their research focuses on the Nigerian GSM telecommunications industry, employing a blend of Principal Component Analysis (PCA), KMeans Clustering, and Logistic Regression to develop a robust model predicting customer churn based on Quality of Service (QoS) metrics from major Mobile Network Operators such as MTN, Airtel, Globacom, and 9mobile. The model impressively achieves a prediction accuracy of 95%, demonstrating the efficacy of combining these algorithms for churn prediction.

However, the study is not without limitations. The reliance on QoS parameters, while useful, may not capture the full spectrum of factors influencing churn, such as pricing strategies, which is also critical to customer retention strategies. Furthermore, the study's focus on a single national context might limit its generalisability to other regions with different competitive dynamics and consumer behaviours. Thus, while the study advances the use of machine learning in churn prediction, it also highlights the need for more comprehensive models that incorporate a wider range of variables and cross-regional analyses.

2.2 Customer Churn Prediction for a Telecommunication Company

In the study by Behl et al.(Behl, Hadpe, Bonde, & Patel, 2024) the authors address the critical issue of customer churn in the telecommunications sector, utilising advanced machine learning techniques to predict churn based on vast volumes of historical customer data, including demographics, usage patterns, and service interactions. The study showcases the development of a robust predictive model that effectively forecasts the likelihood of churn, allowing for proactive customer retention strategies such as personalised incentives and targeted service offerings.

Despite its comprehensive methodology and practical implications, the study acknowledges several challenges related to data privacy, model interpretability, and the integration of predictive analytics into existing workflows. These challenges underscore the complexities of implementing machine learning solutions in a real-world business context, which may hinder the full realisation of such models' potential benefits.

Behl et al. contribute significantly to the understanding of customer churn in telecommunications, providing valuable insights that aid in the development of data-driven strategies to enhance customer retention and competitive advantage.

2.3 Empirical Analysis of Tree-Based Classification Models for Customer Churn Prediction

The study by Usman-Hamza et al.(Usman-Hamza, 2024) provides a comprehensive evaluation of tree-based classification models for customer churn prediction (CCP) in the telecommunications industry. The paper stands out for its in-depth exploration of various tree-based classifiers including decision trees, random forests, and various ensemble and hybrid models, demonstrating their effectiveness in handling class imbalance problems - a common challenge in CCP datasets. The researchers systematically investigate the performance of these models under different conditions, including the presence of class imbalances, and propose methods such as data sampling techniques to improve model accuracy.

However, the study also acknowledges certain limitations. Despite the advanced techniques applied, the generalisability of the results to other sectors or broader datasets might be constrained due to the specific characteristics of the telecommunications data used. Furthermore, while the paper discusses the potential of tree-based models to offer interpretable insights, it also notes the complexity that comes with some of the more sophisticated ensemble and hybrid models, which could complicate their practical deployment without expert interpretation.

Overall, this research significantly contributes to the field by enhancing the understanding of how different tree-based models can be optimized for predicting customer churn, offering valuable insights for both academic researchers and industry practitioners looking to reduce churn through targeted interventions based on predictive analytics.

2.4 Dynamic Churn Prediction System Using Machine Learning Algorithms

The study by Bhadane et al.(Bhadane, Randhir, Borade, Bhatia, & More, 2023) delves into developing a dynamic churn prediction system utilising a machine learning algorithm to tackle the persistent challenge of customer churn in the telecommunications industry. This paper stands out for its integration of real-time data processing with an advanced analytical model which significantly enhances the ability to predict churn promptly and accurately. The system's dynamic nature allows it to adapt to new patterns in customer behaviour, ensuring that the churn prediction remains relevant over time.

However, while the paper presents a robust approach to churn prediction, it does have limitations. The reliance on complex machine learning models requires substantial computational resources and expertise in model tuning, which may not be feasible for all companies. Additionally, the effectiveness of the model depends heavily on the quality and granularity of the data available, which can vary significantly between organisations and markets.

Overall, the research provides a substantial contribution to the field by showcasing how machine learning can be effectively utilised to enhance customer retention strategies in a highly competitive market. By focusing on the proactive identification of at-risk customers, telecom companies can more effectively allocate their customer retention resources and potentially improve overall customer satisfaction and loyalty.

2.5 Machine Learning Based Churn Prediction in Telecom

The paper by Xuhao Dai(Dai, 2023) delves into customer churn prediction within the telecommunications sector, employing machine learning techniques like Logistic Regression, Decision Trees, and Support Vector Machines (SVM) to analyse and predict churn trends based on extensive customer data. The research methodically evaluates the effectiveness of these models, revealing limitations such as SVM's struggles with unbalanced data, as evidenced by specific performance metrics like lower recall and F1-score.

The study also highlights the critical role of integrating diverse data sources and the importance of accurate feature selection to enhance model performance. Challenges in data quality and model complexity are discussed as factors that significantly impact prediction accuracy.

In summary, Dai's research contributes valuable insights into the use of advanced analytics for churn prediction, suggesting that a deeper understanding of customer behaviours, combined with enhanced data integration techniques, is essential for improving model accuracy and effectively reducing customer churn in the telecom industry. Further research is recommended to optimise these machine learning models and refine data handling practices.

2.6 Churn Rate Modeling for Telecommunication Operators Using Data Science Methods

Zatonatska et al.(Zatonatska, Fareniuk, & Shpyrko, 2023) investigate the effectiveness of machine learning models for churn prediction within the telecommunications sector. Their research utilises a variety of models, including logistic regression, Random Forest, SVM, and XGBoost.

The study effectively highlights the potential of machine learning to discern key factors influencing customer retention, such as contract types and payment methods, emphasising their impact on churn rates. However, the research is constrained by its reliance on data from a single telecommunications company, which may limit the generalisability of the findings across different market environments.

This work lays a solid foundation but also underscores the need for broader datasets and more sophisticated handling of methodological challenges in future studies.

2.7 Enhanced Predictive Data Mining Algorithm for Fraud Detection and Churn Behaviour Modelling in Telecommunication Systems

The study conducted by Elechi and Michael(Elechi & Michael, 2023) addresses the development of an enhanced predictive data mining algorithm aimed at improving fraud detection and churn behaviour in telecommunications systems. The primary objective is to curb significant revenue losses caused by fraudulent activities and customer turnover.

Their approach incorporates various advanced data mining techniques. Specifically, the study utilises probabilistic models, a Naive Bayesian model, and neural prediction networks. However, the use of linear discriminant functions is also critical to their methodology. They employ computational analytic modeling to develop an adaptive control strategy for fraud detection, emphasizing the dynamic nature of fraud patterns.

2.8 Enhanced Churn Prediction in the Telecommunication Industry

The paper by Oludele et al.(Oludele, Ben, A.C., S.O., & Seun, 2020) introduces a sophisticated churn prediction model specifically designed for the telecommunication industry. This model is centered around the use of a Markov Chain Model, which provides significant flexibility compared to traditional models by allowing the incorporation of variable retention rates and a probabilistic approach to understanding customer behaviours over time. The model's capability to track individual customer relationships enhances its predictive accuracy and utility.

A key innovation in their methodology is the employment of Monte Carlo simulations within MATLAB. These simulations are utilised to run numerous fictitious scenarios of customer-company interactions, which help in deriving a more accurate estimation of churn rates. This method not only captures the dynamic nature of customer churn but also provides telecom companies with a robust tool to enhance customer retention strategies.

However, the reliance on sophisticated statistical techniques and the need for high-quality, real-time data pose significant challenges. The complexity of the model requires robust data infrastructures and could limit its applicability in environments where data collection is fragmented or analytics capabilities are less developed. Additionally, the indefinite state space of the Markov Chain could complicate the computational demands of the model.

Overall, the study makes a notable contribution to the field by developing a predictive model that not only aids in understanding the immediate factors influencing churn but also enhances strategic decision-making in customer relationship management within the telecommunication sector. The paper serves as a valuable resource for both practitioners in the telecommunications field and researchers interested in predictive analytics and customer behaviour modeling.

2.9 Contribution

This study aims to advance the field of churn prediction in telecommunications by integrating and enhancing machine learning techniques in several pivotal areas.

The projected innovations are expected to not only elevate predictive recall but also broaden the methodological tools available for effectively addressing churn. The contributions include:

- **Hybrid Modeling Approach**: Developing a novel hybrid model combining Gradient Boosting Machines (GBMs) and Random Forests.
- **Model Stacking**: Utilising model stacking to improve prediction accuracy by leveraging multiple models.
- **Comprehensive Model Evaluation**: Employing a range of performance metrics for rigorous model evaluation.
- **Cross-Validation Techniques**: Using advanced cross-validation methods to ensure model robustness and generalisability.

3 Data

3.1 Data

In this section, I will provide an overview of the dataset used for this paper. The dataset contains 12 features that are crucial for understanding customer behaviour and predicting churn in the telecommunications industry. These features are as follows:

CustomerID	A unique identifier for each customer
Age	The age of the customer
Gender	The gender of the customer
Tenure	The number of months the customer has stayed with the company
Usage Frequency	The number of times the customer has used the service in the past month
Support Calls	The number of support calls the customer has made in the past month
Payment Delay	Number of days the customer has delayed payment in the past month
Subscription Type	The type of subscription the customer has
Contract Length	Duration of the contract
Total Spend	The total amount the customer has spent
Last Interaction	Number of days since the last interaction the customer has had with the company
Churn	Whether the customer has churned or not

Table 1: Dataset features

The dataset contains about half a million entries. Having this many entries provides an extensive amount of data for building and evaluating predictive models. This large dataset enables the models to capture a wide variety of customer behaviours and patterns, which is crucial for accurate churn prediction. It allows for robust training, ensuring that the models are not overfitting to a small sample of data but generalising well to new, unseen customers(entries).

However, managing and processing such a sizable dataset also comes with challenges. It requires significant computational resources and efficient data handling techniques. The large volume of data necessitates thorough cleaning and preprocessing to ensure data quality and consistency.
Despite these challenges, the extensive data enhances the reliability and robust-

ness of the predictive models, making the findings more applicable to real-world scenarios.

3.2 Explaining/Predictive Variables

If we take a deeper look into our dataset variables, we find that the dataset can be split into two bigger categories:
Explaining variables and predictive variables, the latter only containing the churn variable. The explaining variables can be further split into subcategories themselves. These include:

- **Categorical values:** Gender, Subscription Type, Contract Length
- **Numerical values:** Age, Tenure, Usage Frequency, Support Calls, Payment Delay, Last Interaction

3.2.1 Categorical Values

Let's take a deeper look at the categorical values. The gender variable helps in analysing if gender has any significant impact on customer churn. For instance, we may observe different churn rates between male and female customers, which can help tailor marketing and retention strategies.

The subscription type variable may influence the extent to which the customer is satisfied. Different subscription types offer varying levels of service and benefits, which may influence churn behaviour.

The contract length variable may also indicate how likely a customer is to churn. A customer with a longer contract length may be less likely to churn due to the commitment required, as opposed to customers with shorter contracts who have more flexibility to leave.

3.2.2 Numerical Values

And, of course, we cannot neglect the numerical values. The age of the customer can influence churn patterns as different age groups may exhibit varying levels of loyalty, technology adoption and service usage. For example, younger customers might churn more frequently in search of better deals.

Longer tenure often indicates higher customer loyalty and satisfaction, which may result in lower churn rates.

Higher usage frequencies generally correlates with higher engagement and satisfaction, reducing the likelihood of churn.

Many support calls in a period may indicate unresolved issues or dissatisfaction, which can increase churn risk.

Customers who delay payments might be experiencing financial difficulties, which may lead to a higher churn risk.

A high value for the last interaction feature may suggest a lack of necessity for the customer, which, in turn, may lead to a higher probability of churn.

The total spend variable indicates the total amount the customer has spent. A higher total spend may indicate greater value derived from the service and potentially lower risk of churn, as said customer may have more invested in staying with the company.

3.3 Descriptive Analysis

In this section I will dive deeper into how these features have correlated with a customer churning. I will be looking at the mean churn rate for each value for an isolated feature and compare them to the overall churn rate. For numerical variables, I have split them into meaningful categories. Let us go through each feature and see if the assumptions I made in the previous section are valid or not. The overall churn rate is 55.5%. Any group that is significantly different from this rate is a group to watch.

3.3.1 Single features

Gender As we can see in the chart, there is no significant difference in churn behaviour between men and women. This aligns with the fact that I did not make any assumptions here. Although customers that have identified as female are slightly more likely to churn(64.9%).

Subscription Type The different churn behaviour among the types are negligible.

Contract Length Here is the first place where we see a significant churn indicator. As opposed to the quarterly and annual subscriptions, customers with a monthly subscription have a churn rate of 90.2%. This confirms my previous suspicions that customers with a monthly subscription are very likely to churn.

Age As opposed to my earlier suspicions, the age group 50+ is more likely to churn with a rate of 83.5%. While the other two groups sway from the overall mean churn rate, customers aged 34 to 49 are slightly less likely to churn with a rate of 42.8%.

Tenure The churn rates here are negligible for all groups (all within 2% range of each other). This shows that the tenure feature may be irrelevant to us.

Usage Frequency Here we see customers who have used the service less than 11 times the past month are slightly more likely to churn (churn rate: 59.8%) than customers who use it more than that. Although the difference is not significant.

Support Calls Now this is more like it! you can see a correlation between the amount of support calls made in the past month and churn rate. The more calls made by the customer, the churn rate rises. The mean churn rate of customers who made between seven and ten calls the past month is 91.5%. The group with four to six calls has a churn rate of 75.2% and the group with less than four calls the past month has a mean churn rate of 32.3%. As is evident, my assumption was correct.

Payment Delay As is shown in the chart, there is a clear spike in churn rate for the customers who delayed their payment for more than 19 days. The churn rate for this group of customers is 94.2%. Customers who delay it less than 20 days show no aggresive churn behaviour. This also confirms my previously stated assumption.

Last Interaction This feature is mildly similar to the Support Call variable, as the churn rate slightly increases for each group topping out at 63.8% for the customers who have not used the service in more than 20 days. This also confirms my previous suspicion.

Total Spend Here again, my assumption was correct, in that customers who have less invested in the company are far more likely to churn compared to customers who have a lot invested, with the percentage of 90.2% for customers who have spent less than 400$ and 41.4% for customers who have spent between 700$ and 1000$.

I will go into further details on the results of the descriptive analysis in the Results chapter.

4 Model Construction

In this chapter I will talk about how I prepared and implemented my predictive model. Now, before we address anything, it is important to note that all the categorical features were transformed into a one hot encoded version. For example instead of having one feature called Gender with possible values being male or female, it is now split into two columns(isMale, isFemale) with binary values. Take Customer A. A is female. Before the one hot encoding A's entry would have a feature called Gender with the value Female. Now it has dropped that column and added two different ones: gender_Male and gender_Female with the values 0 and 1 respectively. The reason why I used the OneHotEncooder instead of a label encoder is simple. The label encoder encodes all categorical values of a feature into an ordinal set of integers. For example, as opposed to the one hot encoding method, the label encoder does not create more columns for each categorical value. It instead converts them into numerical labels. for example, instead of having one column for gender with the values Female and Male, there would still only be one column, but the values would be 0 and 1, each presenting their own categorical value. The issue here is the ordinality of the numbers. The model will then implicitly interpret one of the categorical values to be greater than another. This could easily lead to problems later on when implementing the predictive model.
This is why we use the one hot encoding approach instead of the label encoding approach.
The essence of my thesis is designing a novel approach to predicting customer churn. In order to asses this model's performance, I have implemented a primitive model which I will go into further detail later on. The intention of this model is to use as a baseline to see whether my extra efforts would be worth it.
PS: I feel like it is a given, but I am still going to mention it for the sake of clarification: For most Models there are Regression- and Classification-versions. As we are trying to predict a binary value(Will the customer churn or not), I will be using the classification models exclusively.

4.1 Chosen Models

In designing my model, I went through a plethora of different models and compared them. I ended up with the following models:
Here I will talk about the base models used in my approach.

RandomForest RandomForest is an ensemble learning method used for (in our case) classification tasks. It builds multiple decision trees and merges them together to get a more accurate and stable prediction. Here's a detailed look at how the model works:

- **Bootstrap Aggregating (Bagging):** RandomForest uses a technique called bagging, where it creates multiple subsets of the original dataset. Each subset is used to train a separate decision tree.

- **Random Feature Selection:** When splitting a node in each decision tree, RandomForest randomly selects a subset of features. This ensure that the trees are not highly correlated and increases the overall diversity and robustness of the model.

- **Aggregation of Results:** For classification tasks, each tree in the forest outputs a class prediction, and the class with the most votes becomes the model's prediction.

One of the advantages of the RandomForest model is that, by averaging multiple trees, the model reduces the risk of overfitting compared to individual decision trees. This typically leads to a high predictive accuracy and is very capable in handling larger datasets with higher dimensionality.

Although the model does come with its own drawbacks. For one, you have to be careful and limit the number of trees. Otherwise the model can become computationally expensive. This could lead to slower predictions.

XGBoost XGBoost, which stands for Extreme Gradient Boosting, is an optimised distributed gradient boosting library designed to be highly efficient, flexible, and portable. It implements the gradient boosting decision tree algorithm. The working mechanism is as follows:

- **Boosting:** Unlike bagging, boosting builds trees sequentially. Each tree corrects the errors of its predecessor.

- **Gradient Descent:** XGBoost uses gradient descent to minimise the loss function, to improve the model's performance with each iteration.

- **Regularisation:** It includes regularisation terms (both L1 and L2) to control overfitting and improve generalisation.

Another thing to mention is that the model is capable to handle sparse data efficiently.

Some advantages of this model is that, XGBoost often delivers superior performance in terms of speed and accuracy compared to other models. Although it is able to handle sparse data efficiently, it is designed to scale to an extreme number of examples in distributed or memory-limited settings.

As this model can become quite complex, it is extremely important to perform careful hyperparameter tuning to achieve optimal performance.

LightGBM LightGBM, which stands for Light Gradient Boosting Machine, is another gradient boosting framework that is highly efficient and scalable, specifically designed to handle large-scale data with higher performance. The inner working mechanisms of the framework is as follows:

- **Leaf-wise Growth:** Unlike other boosting algorithms that grow trees level-wise, LightGBM grows trees leaf-wise. This could result in deeper trees and potentially better accuracy.

- **Histogram-based Decision Tree:** It uses a histogram-based approach to find the best split, which significantly reduces memory consumption and speeds up training. This is will also be perfectly visible later on during hyperparameter tuning.

- **Optimisation:** LightGBM includes several optimisations, such as Gradient-based One-Side Sampling and Exclusive Feature Bundling to handle larger datasets efficiently.

One of the biggest advantages of the LightGBM model is that it is extremely fast, often outperformng other gradient boosting frameworks in terms of training speed. For example, training the model 81 times with five cross-validation folds took around 8 minutes compared to XGBoost, which took almost 2 hours (on my machine) to train a model 100 times with 3 cross-validation folds.

LightGBM can also handle large datasets efficiently. The model also consumes less memory compared to other boosting algorithms. It also generally provides predictions with a high performance, especially on large datasets.

Now, it might seem like this model is absolutely amazing. However, as with almost every single model, it doesn't come without its disadvantages. Especially, the model, due to its Leaf-wise growth, can lead to overfitting if not properly controlled. Similarly to XGBoost, the model can be hard to interpret and needs careful hyperparameter tuning to achieve the optimal predictions.

Decision Tree The Decision Tree model is a non-parametric supervised learning method. It creates a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. The inner working mechanisms of the Decision tree is as follows:

- **Tree Structure:** A decision tree is composed of nodes and branches. Each internal node represents a decision based on a feature, each branch represents the outcome of the decision, and each leaf node represents a class label(in our classification case).

- **Splitting Criteria:** The tree uses the Gini impurity criterion as a standard. But it can be set to other criteria as well.

- **Recursive Partitioning:** The process of splitting the dataset continues recursively until a stopping criterion is met, such as a maximum depth or minimum number of samples per leaf.

There are multiple advantages to this model. One being its simplicity and interpretability. It is easy to understand and interpret, as the decision-making process is straightforward. Once trained, the model's predictions are fast and efficient.

However, there are also risks to this simple model. One being the high risk of overfitting, especially with deep trees. Small changes in data can also result in very different tree structures.

As you can probably see, all the chosen models are tree-based models. Some are simple, such as the basic decision tree model, while others are complex ensemble models based on simple decision tree models. This special approach of only using tree-based models comes with its own advantages and disadvantages.

One of these advantages is that tree-based models can handle both numerical and categorical data. They are also able to capture non-linear relationships and interactions between feature, making them suitable for a wide range of classification problems.
Decision trees, even in ensemble forms like RandomForest and to some extent in gradient boosting methods, offer a level of interpretability. They provide insights into the feature importance and decision-making process, which is valuable for understanding the model's behavior and for explaining predictions to stakeholders.

Ensemble methods like RandomForest, XGBoost and LigthGBM leverage multiple trees to improve accuracy and robustness. They reduce the risk of overfitting, which is a big issue seen in single decision trees. This could be done by averaging or boosting, resulting in more stable and reliable models.

Now, on the other hand, Training these complex ensemble models individually are computationally expensive and very time consuming, especially on our large dataset. Even more evident is this during hyperparameter tuning. There are so many hyperparameters to choose from that it doesn't make sense to waste days, or even weeks in some cases (except if you have a supercomputer or have the funds to rent multiple GPUs online), to tune every single one of the hyperparameters. This will be made very clear later on when talking about hyperparameter tuning. Decision trees and their ensembles can also exhibit bias towards features with many levels or higher cardinality, potentially skewing the model's focus. This can affect the model's ability to capture relevant patterns effectively, especially if some important features are overlooked due to their lower cardinality.

This is where the stacking approach comes into play. It works with a stack of base estimators and a final meta classifier. The output of the the individual estimators are stacked and a final estimator is used to compute the final prediction. This allows us to use the strength of each individual estimator by using their output as input of a final estimator. In my implementation you will see that I combine these tree-based models to create a powerful stacking model. This stacking model leverages the strengths of each base model. But more on that later.
Although one thing to look out for when using tree-based models exclusively, is, more than anything, the danger of overfitting.

4.2 Primitive Approach

Before we dive into my approach, let us go through the primitive, baseline approach.
before doing anything, the dataset needs to be prepared.

1. First the variables X and y are set to contain the explaining and the predictive features respectively.

2. In the second step, a simple test and train split is created with a test size of 20%.

3. both of the explaining feature sets are now scaled using a standard scaler.

A standard Ridge classification model is trained on the training set and then, finally, the model tries to predict the churn behaviour on the test set. The results of these predictions will be discussed in the next chapter.

4.3 My Approach

There are many differences between my approach and the baseline approach. Let us first look at the data preparation:

1. Like in the primitive approach, we first set X and y to contain the explaining and predictive data respectively.

2. In my approach I split the data into training, validation, testing sets. This is done in order to have an extra layer of data, so I can also test the results of the hyperparameter tuning.

3. Finally, the features are standardised. This is also done using a standard scaler.

4.3.1 Hyperparameter Tuning

My first step is to perform hyperparameter tuning for the base models. It is important to note: all tuning done here is with the training set. We do not use the validation or testing set yet
There were three main things I had to take into account:

1. Which hyperparameters to be tuned and what parameters should be tested.

2. whether I should continue with a Grid-search or Randomised-search, the main difference is that grid search goes through every possible combination for the previously selected hyperparameters. The Randomised search picks a specified number of combinations and tests them. I will explain why I used which method for every model.

3. what performance metric should be used to score and compare the combinations of hyperparameters.

RandomForest I began with the hyperparameter tuning of the RandomForest model. For this I used the following hyperparameters: Whether Bootstrap samples are used when building trees, the maximum depth of the tree, the minimum number of samples required to be a leaf node, the minimum number of samples required to split an internal node and the number of trees in the forest.
As I chose a lot of possible values for some of the hyperparameters, it made sense to use the randomised search method, as it would take too long otherwise. The

disadvantage of this is, I might not find the optimal hyperparameters, but with some luck I can find a good enough one.

I decided to search for 100 random candidates and, for each of those candidates. I fitted 3 folds, totalling 300 fits.
I used recall for all the models, as I believe this is the most important performance metric. This is because it is more important to correctly classify true values than anything else (I will go into further detail in chapter 5).
The best model will be saved for further use.

XGBoost The second model to be tuned was XGBoost. First we will go through the different hyperparameters I chose to tune:

- The step size shrinkage used in update to prevent overfitting
- The number of boosting rounds, which is the total number of trees to be built
- The maximum depth of a tree, which controls the complexity of the individual trees in the model
- The minimum sum of instance weight needed in a child
- The subsample ratio of columns when constructing each tree
- The subsample ratio of the training instance. This controls the number of rows used to grow each tree

Just like the randomForest model, there are a lot of possible combinations with the chosen hyperparameters and their values, so it would make more sense to use the randomised search method. However, as the XGBoost model is on the faster side, we are able to fit more samples than with the randomforest model. I decided to go with 300 possible candidates, which were then fitted on each of their three folds, totalling 900 fits. As for every model, I decided to go with the recall metric as the scoring system.

LightGBM Now for the LightGBM model, I decided to cut down on the amount of hyperparameters, and their respective values, to be tuned. This allowed me to use the grid search method and instead of three folds for the cross-validation, I decided to go with five. The hyperparameters that I tuned for the model are:

- The maximum depth of the trees,

- The contribution rate of each tree to the final model,
- The number of leaves in a tree,
- And the number of boosting iterations(# of trees in the model)

As I do not have that many possible combination, and on account of the LightGBM model's speed, I decided to go with the grid search method to go over each of the possible candidates and fitted them on 5 folds. This resulted in 405 total fits. Just like with all the other models, I decided to use the recall metric as the comparison score for the model.

Decision Tree The last model to tune is the Decision tree model. This model is also very fast to train, so if we do not overdo it with the number of possible combination it is feasible to do a complete search of all the possible candidates and fit them over 5 folds. The hyperparameters are as follows:

- The maximum depth of the tree,
- The minimum number of samples required to split an internal node,
- The minimum number of samples required to be a leaf node,
- And the number of features to consider when looking for the best split

With these hyperparameters it comes out to 144 possible candidates. When taking cross-validation into account it is 720 total fits. And again we use the recall metric to score the candidates.

Now, that we have the best hyperparameters for the most optimal models, we can go forward to the next step.

4.3.2 Implementing the stacking model

For my ensemble learning technique I used the stacking classifier. This combined my base classification models and created a more robust and accurate final model. It works by taking the predictions of the base models and uses them as features. These features are trained to create a new prediction. This is the meta model that needs to be declared as well. Usually, a primitive model like a linear regression (for a regression model). However, in our case, a primitive model proved to be ineffective (worse than the best base model), so I decided to go with the best base model as a meta model. This was the lightGBM model. As it turns out, this made it much better.

Now, the big question is, should we perform hyperparameter tuning on this model? As a Stacking classifier, it took significantly longer to fit it than the base models (about 1.5 hours on my machine). Now, of course, you could use online resources (paid) to boost your computational power. I decided not to do this as the model performed very well without its hyperparameters tuned (probably on account of the base models being individually tuned).

Although, in our case, the meta model (LightGBM classifier) was also used as a baseline model, for which we also have tuned the hyperparameters. I implemented both stacking models, with and without the best hyperparameters for the base model.
The resulting stacking model, alongside all the base models (for reference), was used to predict the validation set. The results of these tests, and the final test on the test set, will be discussed in the next chapter.

The repository with the jupyter notebook can be found here.

5 Results

In this chapter we will discuss the results of both the descriptive, and the predictive analysis. Although we already talked about the descriptive results a bit in the Data chapter, I will go into further detail here.

5.1 Descriptive Results

I created a correlation heatmap, which visualises the correlation coefficients between different pairs of features in our dataset. Here are some key insights:

Support Calls & Churn Customers who made support calls were more likely to churn, as indicated by a moderate positive correlation 0.52. This suggests that frequent interactions with support may be a sign of dissatisfaction or unresolved issues, leading customers to eventually leave the service.

Total Spend & Churn There is a moderate negative correlation of -0.37 between total spend and churn, meaning that customers who spent more money were less likely to churn. This implies that higher spending is associated with greater satisfaction or a stronger commitment to the service, which helps retain customers.

Payment Delay & Churn Customers who delayed payments were more likely to churn, as shown by a moderate positive correlation of 0.33. Payment delays may indicate the customer is unable to pay for the (possibly expensive) service or a dissatisfaction with the service in itself.

Age & Churn The weak positive correlation of 0.19 between age and churn signifies that older customers were slightly more likely to churn. This trend might reflect changing needs or preferences among older customers, leading them to switch to services that better meet their new requirements.

Usage Frequency & Churn There is a weak negative correlation of -0.053 between usage frequency and churn. This might suggest that customers who used the service more frequently were slightly less likely to churn. This could imply that higher usage may indicate greater satisfaction or necessity of the service, which could very well lead to customer retention. This correlation is almost negligible though.

Support Calls & Payment Delay Customers who made more support calls were somewhat more likely to delay their payments, as indicated by a weak positive correlation of 0.18. This relationship suggests that the issues prompting support calls may also cause financial dissatisfaction or difficulties, leading to delayed payments.

Support Calls & Total Spend There is a weak negative correlation of -0.21 between support calls and total spend. This may indicate that customers who make more support calls tend to pay less money. This could imply that issues requiring support were preventing customers from fully utilising or investing in the service, resulting in lower spending.

Customers who frequently contact support were more likely to churn, possibly due to unresolved issues or dissatisfaction. Higher spending correlates with lower churn, suggesting that more invested customers are more likely to stay. Payment delays are associated with higher churn, indicating financial or service-related dissatisfaction. Older customers are slightly more likely to churn, potentially due to changing needs.
Now, I believe it is important to note: correlation does not prove causation, it is only how past cases have behaved.

5.1.1 Cross-Tabulation

Now, I will go into detail on the cross-tabulation results you can find here. I will be focusing on the combined influence of different feature values on churn rates. By doing this, we can can find nuanced patterns that single-feature analysis might overlook.

Age & Gender The relationship between age and gender reveals distinct patterns in churn rates. Younger customers (aged 18 to 33) exhibit a significant difference in churn rates based on gender. Customers who identify as female in this age group have a higher churn rate (61.9741%) compared to customers who identify as male (44.1193%). Now, in the older age group (50+), while they have a much higher overall churn rate, they do not differ as much, with 86.5439% for customers who identify as female and 80.5493% for customers who identify as male. This could mean that the churn reasons for the younger group differs on the basis of gender identity, while in the older age group, it could mean that they are leaving for the same reasons.

Usage Frequency & Support Calls The interaction between usage frequency and support calls provides valuable insights into customer behaviour. Customers with low usage frequency and high support call frequency have an extremely high churn rate of 92.6775%. This could be on account of customers being so dissatisfied with the service (assumption based on the many support calls) that they stop using the service as much and, in the end, end up churning. On the other hand, customers with a high usage frequency and low amount of support calls only have a churn rate of about 30%. This could suggest better satisfaction and engagement with the service.

Payment Delay & Contract Length The combination of of payment delay and contract length signifies important customer behaviour. It shows similar behaviour among the different contract lengths for both of the lower payment delay groups. With the highest churn rate for monthly customers, showing that the higher degree of flexibility allows the customer to leave the service more freely. However, for customers in the highest payment delay group (20+) have a very similar churn rate of about 93% for annual and quarterly customers, and almost 96% for monthly customers. This shows that payment delay, after a certain point, is a much more dominant feature. This could signify the importance of the payment delay feature when looking at overall churn rate.

Gender & Last Interaction The combination of gender and last interaction sheds light on customer retention and engagement. Customers that identify as female have differing churn behaviour for the different groups. For the subgroup that has last interacted with the service within 10 days the churn rate is 49.9979%, customers that have last interacted with the service within the last 11 to 20 days have an average churn rate of 64.8786%, and customers that have not interacted with the service for more than 20 days have an average churn rate of 90.5873%. This shows that customers who identify as female have a higher and higher churn rate the longer they have not interacted with the service. Now, when looking at the customers who identify as male, we do not see this differing behaviour. Regardless of their last interaction with the service, they have a very similar churn rate of about 48%. This highlights even more the difference between the customers who identify as female and male, with the former taking last interaction into consideration, and the latter not doing so. Or so it would seem.

Tenure & Total Spend Examining tenure and total spend reveals that customers with shorter tenures and lower total spend have a higher churn rate of 88.6392%. On the other hand, customers with the longest tenures and the highest total spend are the least likely to churn with a rate of 40.1643%. This suggests that both longer tenure and higher spending are strong indicators of customer loyalty and reduced churn risk.

Support Calls & Total Spend The interaction between Support calls and total spend shows, like the previously mentioned combination of features, there is a negative correlation between total spend and churn. However, the degree of correlation varies depending on the amount of support calls made in the past month. There are big differences in churn rates for the subgroup of customers who made less than four support calls in the past month with the highest being 87.9785% and the lowest 19.5594%. For the group that made between 4 and 6 support calls in the past month the difference is not as big with the highest being 90.6824% and the lowest 65.5820%. For the last subgroup that made more than six calls in the past month the churn rate is very similar at about 91%. While higher spending generally reduces the likelihood of churn, this effect is moderated by the frequency of support calls. Customers who need more frequent support are at a higher risk of churning, and this risk becomes more uniform with an increasing number of support calls, potentially overshadowing the positive effects of higher spending.

The cross-tabulation analysis highlights the complex interplay between different features and their combined influence on churn rates. By examining pairs of features, we gain a deeper understanding of how various factors interact to affect customer behaviour. These insights can inform targeted strategies to improve customer retention and reduce churn in the telecommunications industry.

5.2 Predictive Results

Even though the earlier results were important, we have now arrived to the heart of the paper, the results of the predictive models.
First, let us talk about my chosen performance metrics and their relative importance. In order to correctly find the right performance metrics, we need to figure out what is important to our use case. The most important thing is to correctly predict true labels so that we can minimise the amount of customers that are going to churn that are not correctly classified. This has to be the absolute priority. The second priority is to minimise the amount of false positives (Customers who are not likely to churn but are classified as churners). Out of this we can extrapolate the best performance metrics for our needs.

1. Recall Recall is our most important metric. Recall measures the proportion of actual positives that are correctly identified. High recall indicates that most of the actual positive cases are being captured by the model.
Recall is calculated like this:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{1}$$

2. Precision Precision is our second most important metric. Precision measures the proportion of predicted positives that are actually correct. After recall, precision is crucial for ensuring our model is as good as possible.
Precision is calculated like this:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{2}$$

3. F1 Score The F1 score is the harmonic mean of precision and recall, giving a balance between the two. In situations where both correctly identifying true labels and minimising false labels are important, the F1 score provides a single metric that considers both.
F1 score is calculated like this:

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

4. ROC AUC ROC AUC provides a measure of how well the model distinguishes between classes across all thresholds. It balances the trade-off between sensitivity (recall) and specificity (true negative rate). While not as direct as recall or precision, it provides an overall performance measure of the classifier's ability to discriminate between positive and negative classes. It can take values from 0 to 1
A higher ROC AUC indicates better performance. A perfect model would have an AUC (Area under the Curve) of 1, while a random model would have an AUC of 0.5.

5. Accuracy Accuracy measures the proportion of all correct predictions among the total number of cases. However, in cases where correct classification of true labels is critical (which it is in our case), accuracy can be misleading. This is why it is of lower importance.
Accuracy is calculated like this:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Number of Cases}} \tag{4}$$

Below is a table, which classifies the performance metrics importances. Of course, in our case, the absolute metric is recall (which is why I used it as the scoring metric in the hyperparameter tuning). I will only talk about recall, precision and F1 score, the remaining metrics can be found in the appendix.

Importance	**Metric**
Highest	Recall
Higher	Precision, F1 Score
Medium	ROC AUC
Lower	Accuracy

Table 2: Metric importance

Now that we have discussed our chosen performance metrics, we can finally delve into our models' performances.
First, we will talk about our baseline results. The baseline model has the following performance metrics:

- Recall: 82.0392%
- Precision: 88.2273%
- F1 Score: 85.0208%

Just as a reminder: There is no validation set here, so the performance metrics here are the only we have for the baseline model. Our goal is to implement a final ensemble model that significantly outperforms this baseline model. Let us first look at the hyperparameter tuned base models (different from baseline model) and their performances on the validation set.

- **RandomForest**
  - Recall: 99.7247%
  - Precision: 89.6704%
  - F1 Score: 94.4307%
- **XGBoost**
  - Recall: 99.7480%
  - Precision: 89.6668%
  - F1 Score: 94.4391%
- **LightGBM**
  - Recall: 99.8945%

  - Precision: 89.6559%
  - F1 Score: 94.4987%

- **Decision Tree**
  - Recall: 99.1867%
  - Precision: 89.6679%
  - F1 Score: 94.1874%

Let us go through each metric one by one and see which is best. But first, we will compare them as a whole to the baseline model. The models all greatly outperform the baseline model in both recall and f1 score, although only slightly outperform it in precision. This is completely fine as 2 out of 3 metrics (4 out of 5 when taking all metrics into consideration) significantly outperform the baseline model. Already here we can see that our more advanced hyperparameter tuned models outperform the baseline model, showing great potential for the ensemble model.
Now lets compare the base models to each other. The best model is LightGBM, with 2 out of 3 metrics outperforming the other ones. The second best is XG-Boost, with RandomForest following in third place and Decision Tree coming in last. Now, we have to mention that the models all perform very well with a recall over 99%, but we want to see if we can improve, if ever so slightly, on the models by stacking them.
The stacking model has the following metrics on the validation set:

- Recall: 99.9732%
- Precision: 89.6316%
- F1 Score: 94.5204%

You might say the improvement from the base models is negligible, However you would be wrong. This is made evident when you look at the confusion matrices of the models' predictions on the validation set. Let us look at the amount of incorrectly classified true labels across the models predicting on the validation set.

- **Random Forest** 159
- **LightGBM** 59
- **XGboost** 141
- **Decision Tree** 450

- **Stacked Model** 15

The difference between the best base model and the stacked model is 44. 44 customers that, if the stacked model was not utilised, would have left the company without it having a chance to retain them. With 15 true labels incorrectly classified out of a total of about 100,000 cases, the model is performing extremely well and we could stop here knowing we have implemented a great prediction model. However, I decided to check one last thing: Will the stacked model perform better with the best hyperparameters for the base model (calculated earlier, when hyperparameter tuning the base LightGBM model). Let us look at how they compare on the test set (the second stacked model is also fitted on the training set).

- **Original Stacked Model**
  - Recall: 99.9775%
  - Precision: 89.6091%
  - F1 Score: 94.5098%
  - Incorrectly classified true labels: 19 out of 151,562 total cases
- **Modified Stacked Model**
  - Recall: 99.9763%
  - Precision: 89.6005%
  - F1 Score: 94.5044%
  - Incorrectly classified true labels: 20 out of 151,562 total cases

As you can see, the original model is performing slightly better. The difference may be negligible though, as you can see by the, compared to the total number of cases, minimal difference between the models. Although, on account of the fact that the original model is still performing better, this will be the final model.

6 Discussion

6.1 Detailed Analysis of Key Attributes

To fully round out our paper, we need to see which attributes influence churn behaviour.

- **Age:** Customers aged 50+ have the highest churn rate of 83.5%. This indicates the need for age-specific retention strategies.
- **Gender:** Customers who identify as female have a slightly higher churn rate (64.9%) compared to customers who identify as male, suggesting the necessity for gender-tailored customer retention approaches.
- **Contract Length:** Customers with monthly subscriptions have a much higher churn rate (90.2%) compared to the customers with annual and quarterly subscriptions, showing the need to further incentivise customers to not sign up for the monthly billing cycle.
- **Last Interaction:** Longer periods since the last interaction are associated with higher churn rates. In order to retain more customers, the service should find a way to keep customer engagement high.
- **Payment Delay:** A sharp increase in churn rate (94.2%) is observed for customers delaying payment more than 19 days. Payment delays signal potential dissatisfaction or financial difficulties.
- **Support Calls:** There is a strong correlation between the number of support calls and churn rate. This shows a need to improve the support experience, to reduce the number of support calls being made.
- **Total Spend:** There is an inverse relationship between total spend and churn, implying the more a customer has invested in the company, the less likely they will be to leave. In order to keep them long enough to not leave before having invested enough, newer customers need to be compelled to stay with the company.

6.2 Implications for Business

6.2.1 Targeted Retention Strategies

Developing personalised retention strategies based on the identified significant attributes is crucial for reducing churn. For instance, the study highlighted that customers aged 50 and above exhibit higher churn rates. Therefore, telecom

companies could offer targeted incentives, such as discounts on long-term plans or exclusive benefits, to this high-risk age group to encourage retention.
Additionally, frequent support callers often indicate unresolved issues or dissatisfaction. By proactively reaching out to these customers with tailored solutions or special offers, companies can address their concerns and reduce the likelihood of churn.

6.2.2 Service Improvement

Improving customer support and reducing payment delays are essential actions for mitigating churn. As mentioned earlier, there is a high correlation between the amount of support calls and churn rate. To address this, telecom companies should enhance their customer support experience by training staff to handle queries more effectively and resolving issues promptly. Implementing a robust feedback system to track and address recurring problems can also help improve customer satisfaction.
Furthermore, long payment delays are a significant predictor of churn. Companies should introduce flexible payment options and send timely reminders to customers, reducing the financial strain and encouraging timely payments.

6.2.3 Product Customisation

Creating customised product offerings for different customer segments is another effective strategy for reducing churn. The study revealed that customer preferences and behaviours vary significantly. For example, younger customers may prefer more flexible, data-heavy plans, while older customers might value reliability and customer service more.
By analysing these patterns, telecom companies can develop tailored products that meet the specific needs of each segment, enhancing customer satisfaction and loyalty. Personalised marketing campaigns that highlight the benefits of these customised offers can further reinforce the value of the service and encourage customers to remain with the company.

6.3 Future Research Directions

6.3.1 Broader Dataset Inclusion

To enhance the model's generalisability across different regions and market conditions, future research should include more diverse datasets. By incorporating data from various geographic locations, market environments, and customer demographics, researchers can develop more robust models that better predict churn across a wide range of contexts. This approach will ensure that the findings are not limited to a specific dataset and can be applied universally.

6.3.2 Feature Engineering

Further exploration into feature engineering is essential to uncover additional significant predictors of churn. By creating new features or transforming existing ones, researchers can gain deeper insights into customer behaviour and improve model performance.
Continuous refinement of features could lead to more accurate and actionable predictions.

6.3.3 Advanced Models

Testing more advanced machine learning models and techniques is recommended to further improve model performance and robustness. While the current model utilises several effective models, exploring newer and more sophisticated algorithms could yield even better results. This could aid in the development of more effective retention strategies.

6.4 Summary of Work Done

6.4.1 Data Preparation and Analysis

The first thing I needed to do was data preparation and analysis. This included several key steps:

Data Cleaning Ensuring the dataset was free of errors and inconsistencies, such as missing values, duplicates and outliers. this step was crucial for maintaining integrity and reliability of the subsequent analysis.

Exploratory Data Analysis Conducting an in-depth examination of the dataset to uncover patterns, trends and relationships among the features. This helps in understanding the distribution of data and correlations between different pairs of variables, providing valuable insights for marketing strategies.

6.4.2 Model Construction

In the model construction phase, various approaches were utilised to build robust predictive models.

Choice of Models Several machine learning models were considered. Each model was selected based on its suitability for handling the specific characteristics of the dataset.

Hyperparameter Tuning This involved optimising the model parameters to enhance performance. Techniques like grid search and random search were employed to find the best combination of hyperparameters for each model.

Implementation of Stacking Models An ensemble method was implemented to combine the base models, leveraging their strengths to create a higher performing and robust final model. The stacking model used the outputs of the base models as inputs for a meta model (in our case, the lightGBM model), which provided the final prediction.

6.4.3 Evaluation Metrics

To evaluate the performance of the models, several key metrics were used:

- **Recall** High recall was crucial for ensuring that most of the churn cases were correctly captured.
- **Precision** Precision was important for minimising false positives.
- **F1 Score** The F1 score, being the harmonic mean of precision and recall, provided a balanced evaluation of the model's accuracy and robustness.
- **ROC AUC** This metric provides an overall performance measure of the model's ability to discriminate between churn and non-churn cases, balancing sensitivity and specificity.
- **Accuracy** While not as critical as recall in our case, accuracy provided a straightforward assessment of the model's overall performance.

These metrics collectively ensured a comprehensive evaluation of the models, focusing on both their ability to correctly identify churn cases and their robustness against false positives.

6.5 Potential Improvements

For brevity's sake, I will not rehash the points mentioned in the Future Research Directions subchapter.

6.5.1 Additional Models

Exploring additional machine learning models and comparing their performance to the current models is recommended. This includes investigating newer algorithms and hybrid approaches that might offer better model performance and robustness. By continuously testing and integrating various models, researchers can identify the most effective techniques for churn prediction.

6.5.2 Continuous Model Updating

Emphasising the need for continuous model updating and maintenance is essential to adapt to changing customer behaviours and market conditions. Regularly retraining models with new data ensures that they remain accurate and relevant. This ongoing process helps in capturing evolving trends and maintaining the effectiveness of the churn prediction system.

6.5.3 Automated Hyperparameter Tuning

Utilising automated tools for hyperparameter tuning (available with scikit learn framework) can save time and improve efficiency. Automated tuning not only enhances model performance but also frees up resources for other critical tasks in the model development pipeline.

6.6 Conclusion

6.6.1 Key Findings

This study has identified several key attributes that significantly influence customer churn in the telecommunications industry. Attributes such as age, gender, usage frequency, support calls, payment delay and total spend were found to be strong predictors of churn. Specifically, older customers and those with frequent support calls or payment delays showed higher churn rates.
Additionally, lower usage frequency and a low total spend were associated with increased churn likelihood.

The predictive models utilised in this study, including RandomForest, XGBoost, LightGBM and Decision Tree demonstrated high effectiveness in identifying churn. Among these, the stacking model, which combined multiple base models, achieved the highest performance, with a recall of 99.9775% on the test set.

These findings underscore the importance of targeted retention strategies and service improvements based on the identified predictors. For example, offering incentives to high-risk age groups or frequent support callers and improving customer support can significantly reduce churn rates.

6.6.2 Impact of Study

The potential impact of this study on business practices, particularly in the telecommunications industry, is substantial. By providing a robust framework for predicting customer churn, this research enables telecom companies to implement more effective retention strategies. Businesses can leverage the insights gained from this study to develop personalised approaches that address the specific needs and behaviours of their customers, thereby enhancing customer satisfaction and loyalty.

Furthermore, this study contributes to existing research by integrating multiple machine learning models and advanced techniques to improve churn prediction performance. The findings and methodologies presented can serve as a valuable resource for future research and practical applications in various industries facing similar challenges.

6.6.3 Future Outlook

Looking forward, the future of churn prediction research should focus on continuous innovation and adaptation to new challenges. The integration of more diverse datasets, advanced feature engineering, and the exploration of sophisticated machine learning algorithms will be crucial for further improving model performance and generalisability.

Encouraging the continuous integration of new data and advanced techniques will enhance the performance and applicability of churn prediction models. By staying at the forefront of technological advancements and data analytics, businesses can maintain a competitive edge in customer retention and overall market performance.

A Appendix

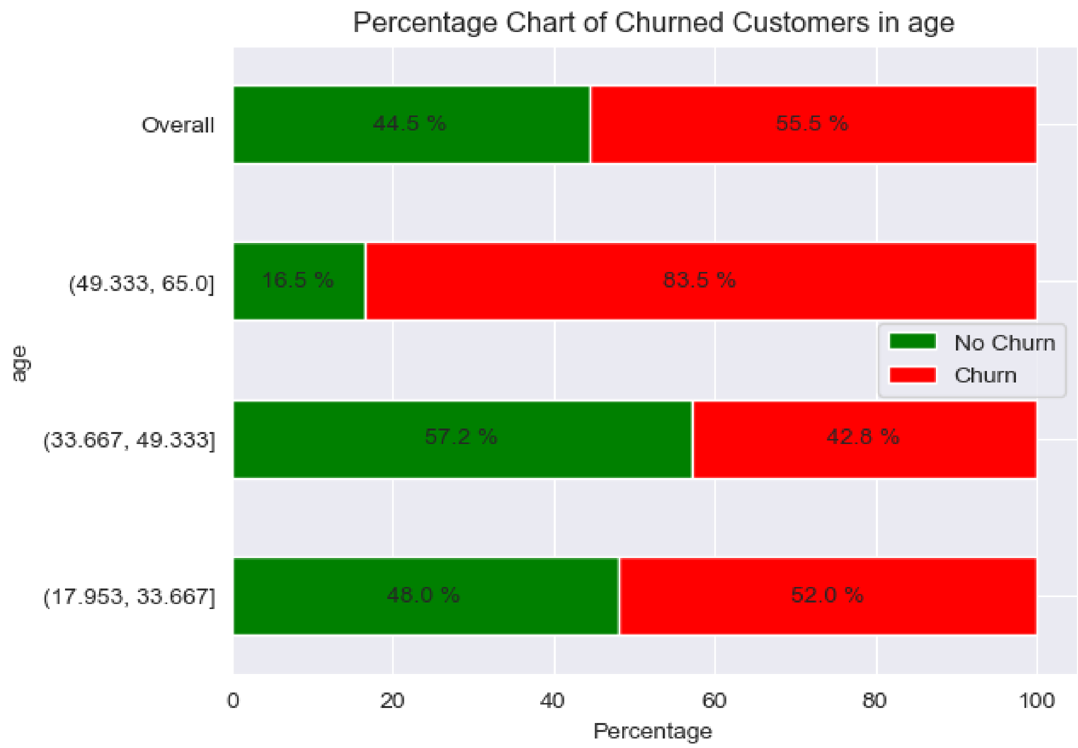

Figure 1: Churn rate for age feature

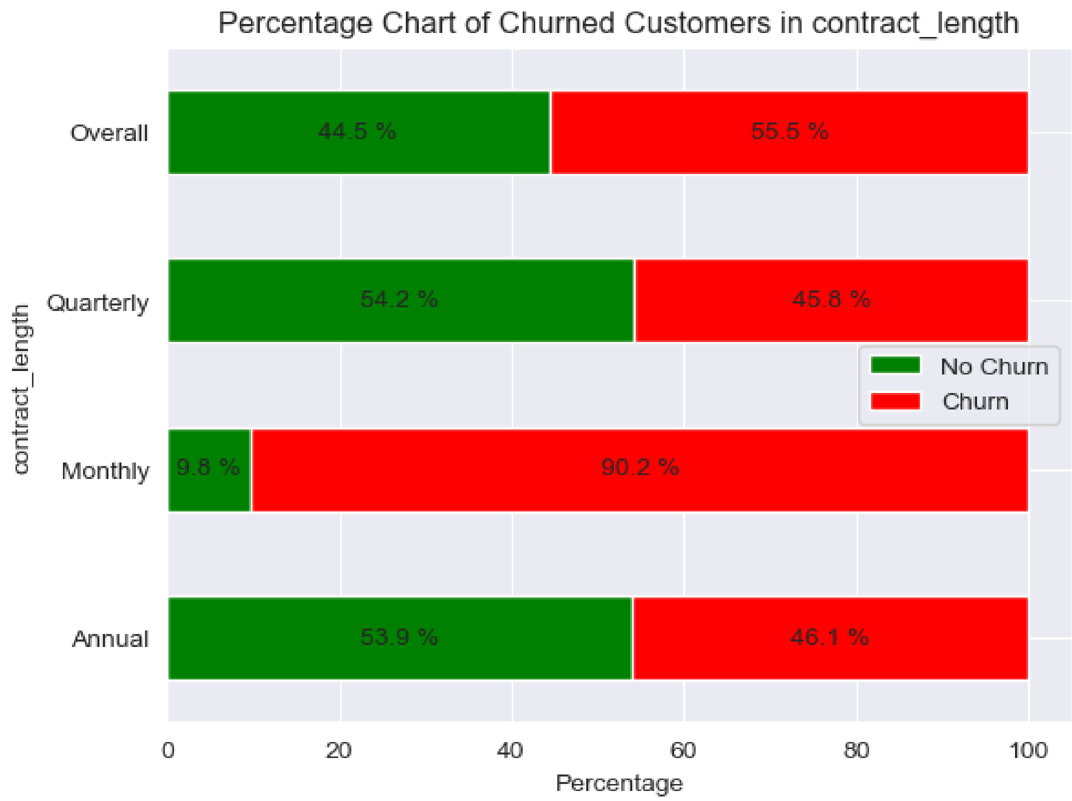

Figure 2: Churn rate for contract length feature

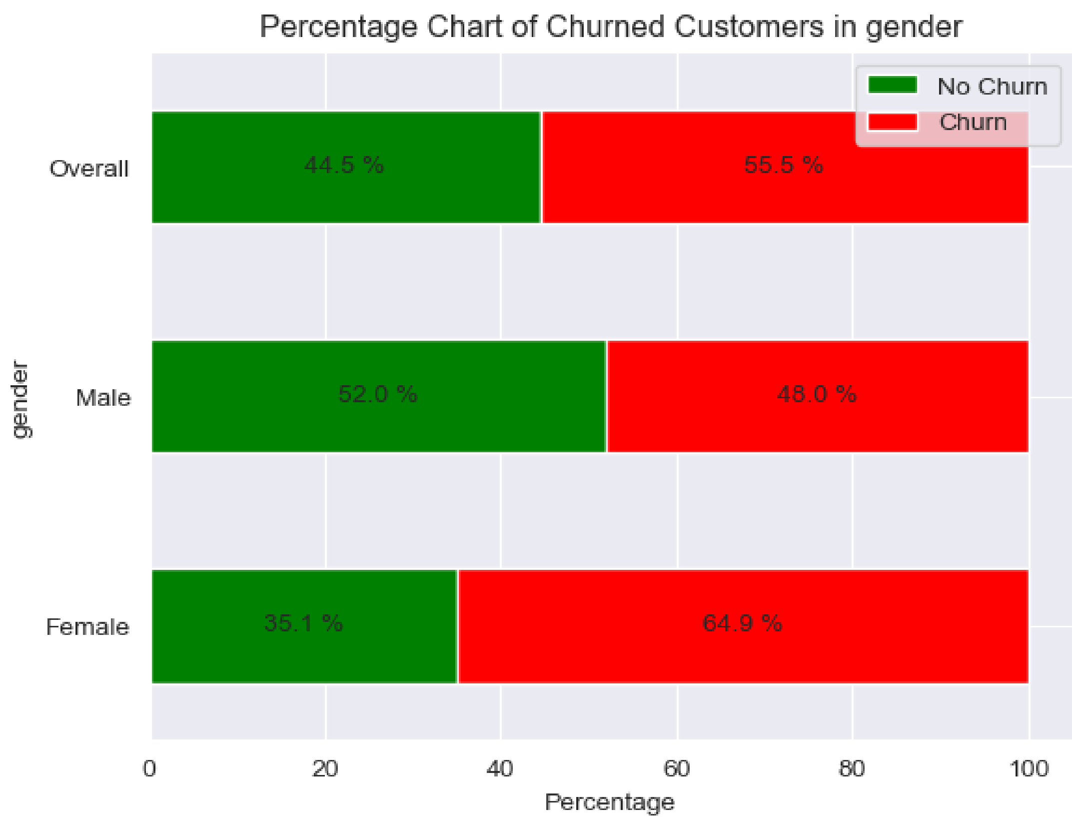

Figure 3: Churn rate for gender feature

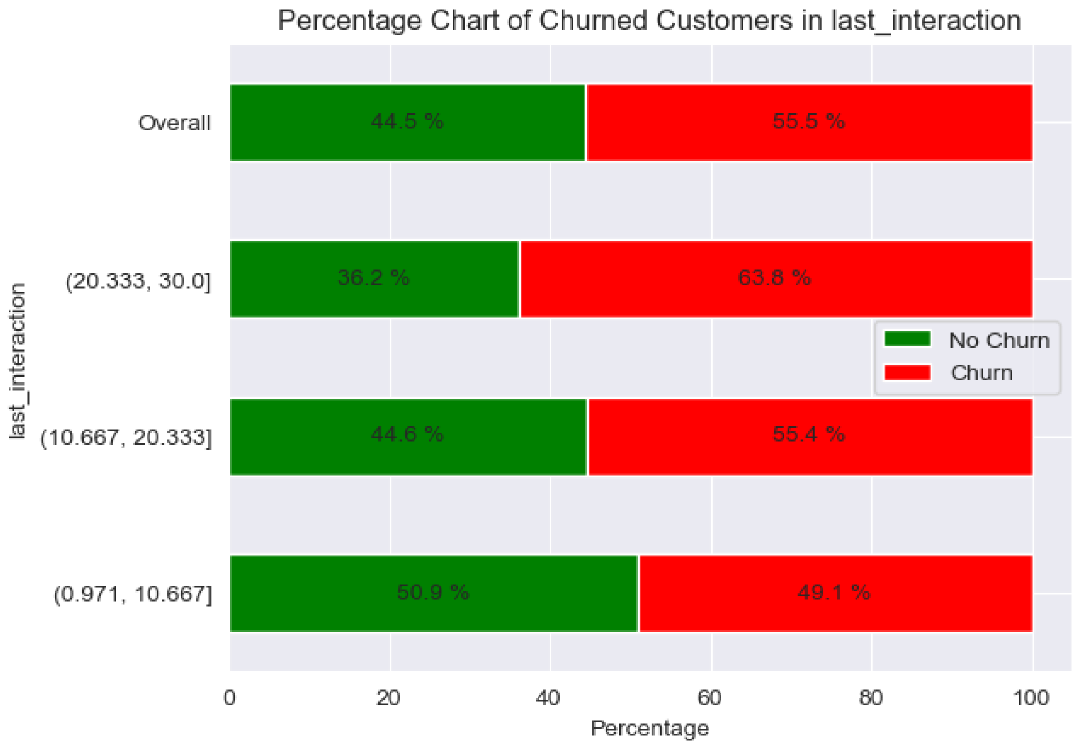

Figure 4: Churn rate for last interaction feature

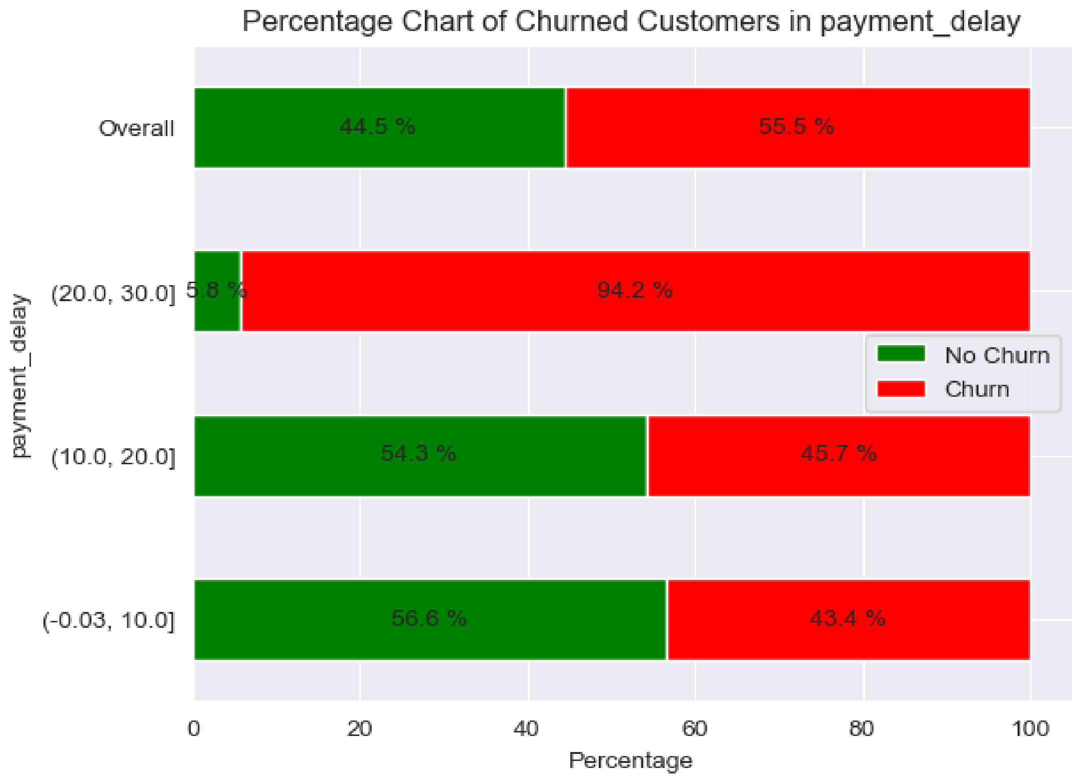

Figure 5: Churn rate for payment delay feature

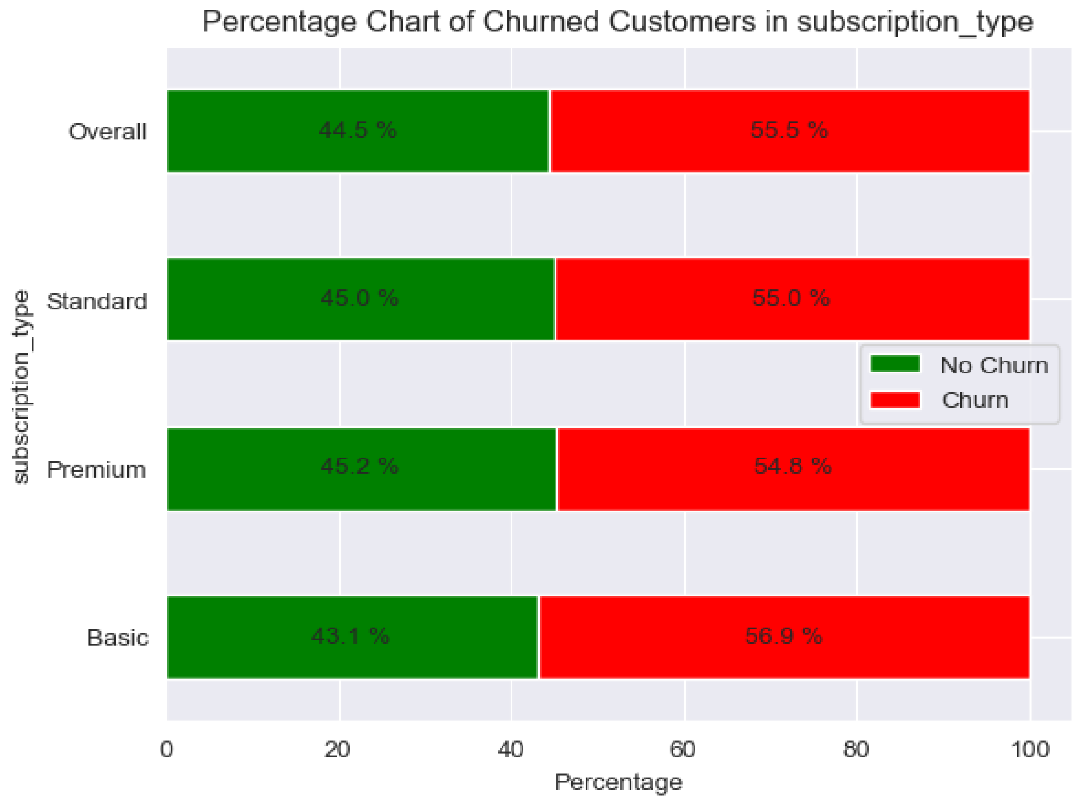

Figure 6: Churn rate for subscription type feature

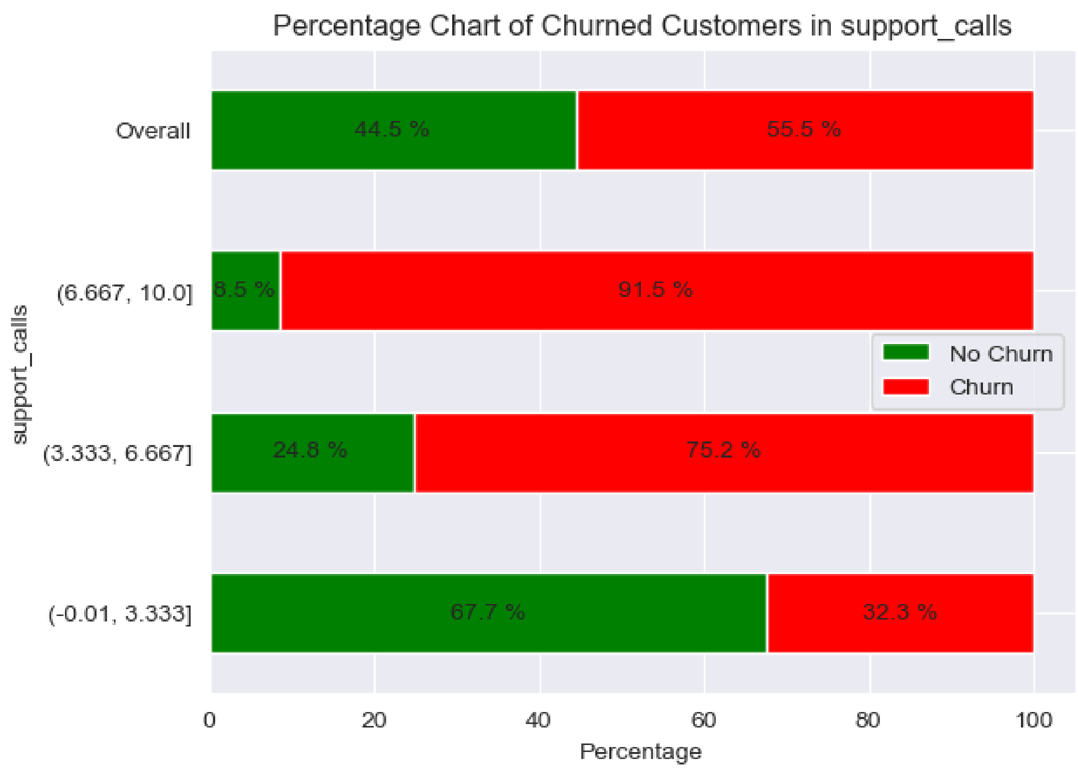

Figure 7: Churn rate for support calls feature

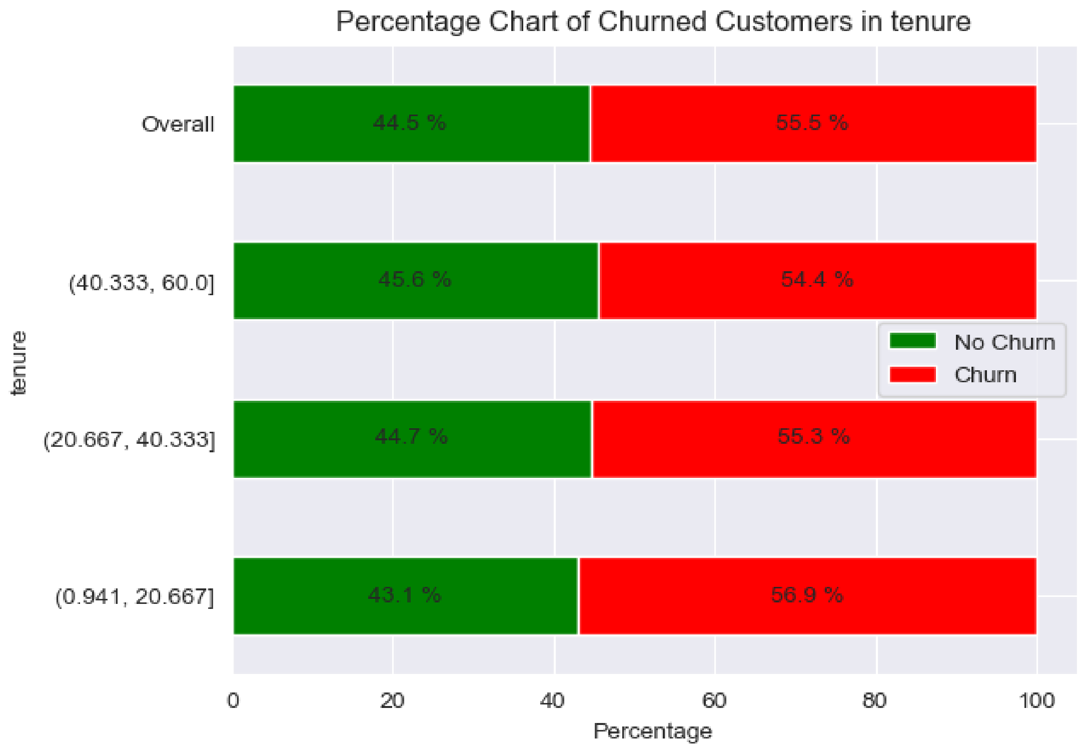

Figure 8: Churn rate for tenure feature

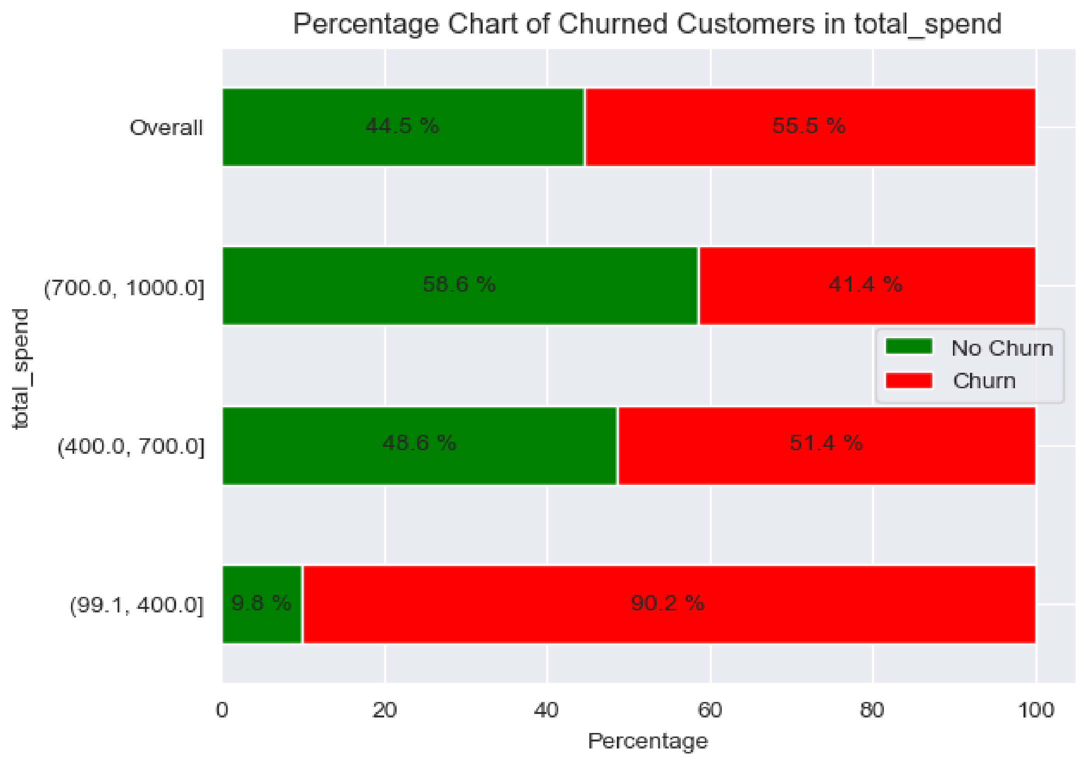

Figure 9: Churn rate for total spend feature

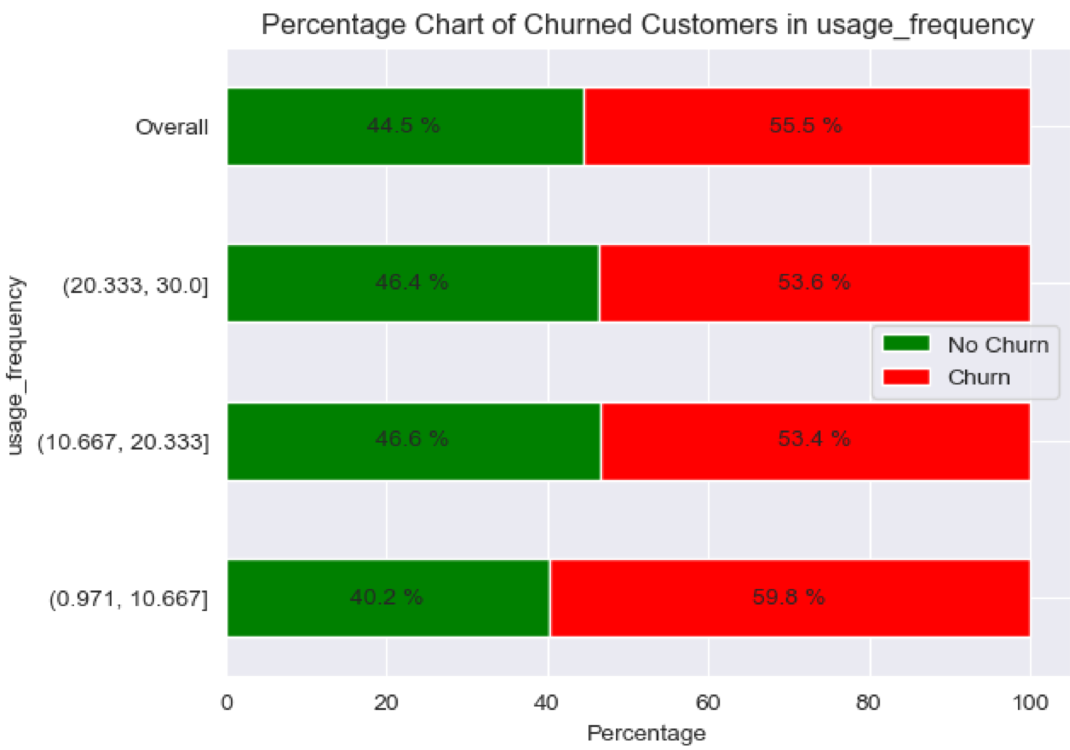

Figure 10: Churn rate for usage frequency feature

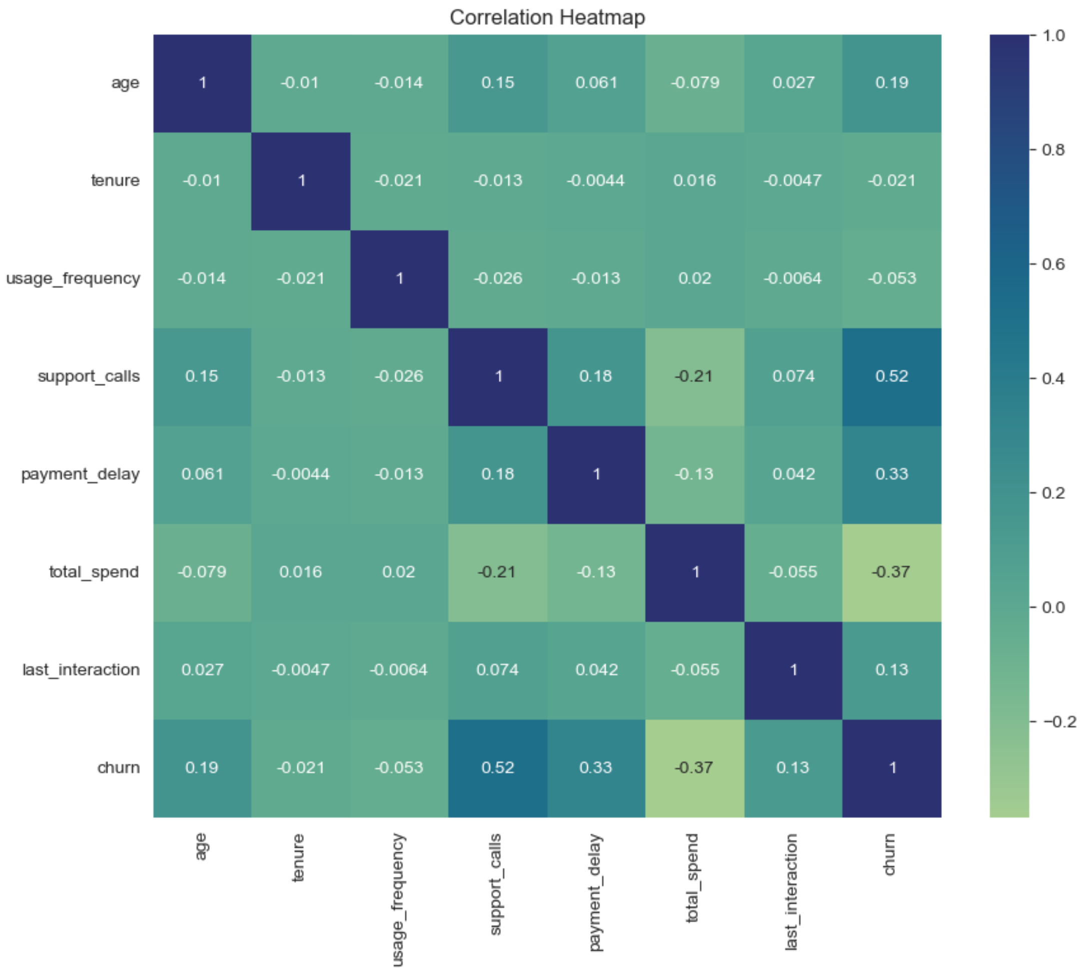

Figure 11: Correlation Map for the features

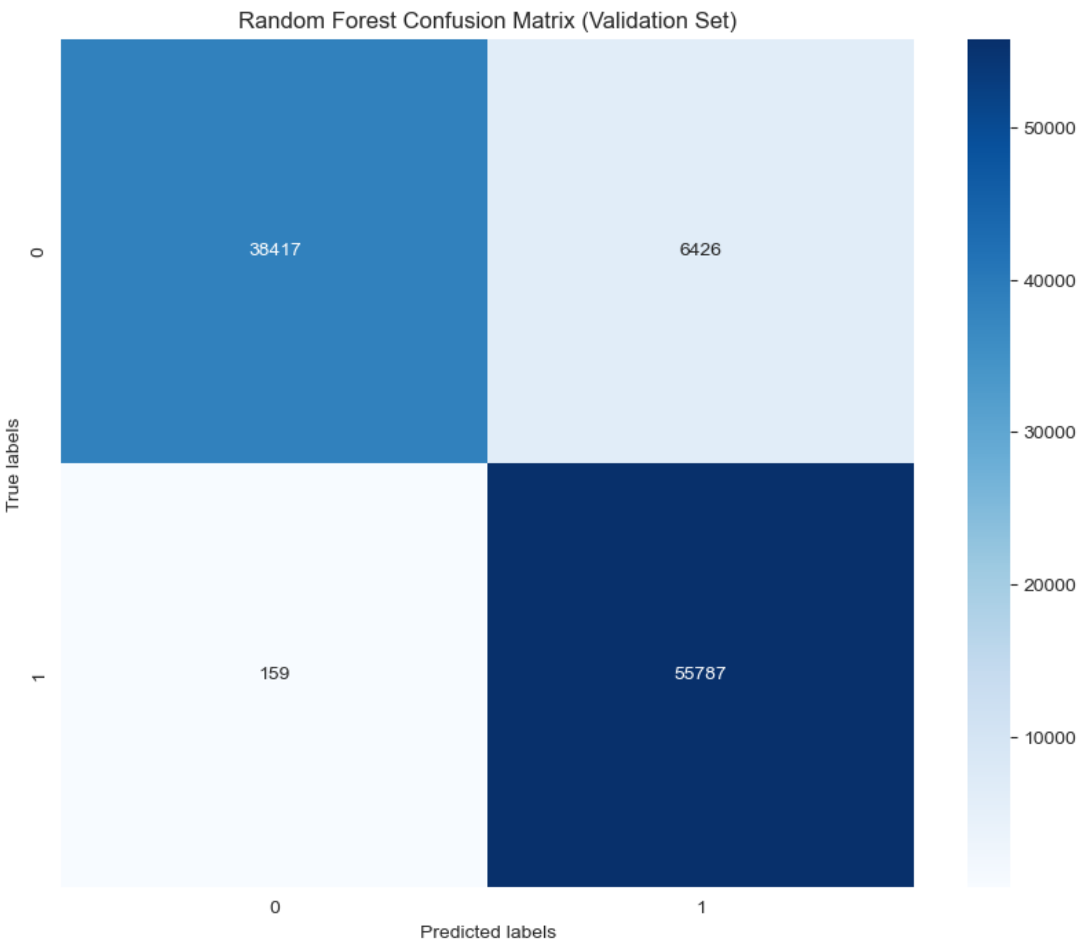

Figure 12: RandomForest Confusion Matrix on Validation Set

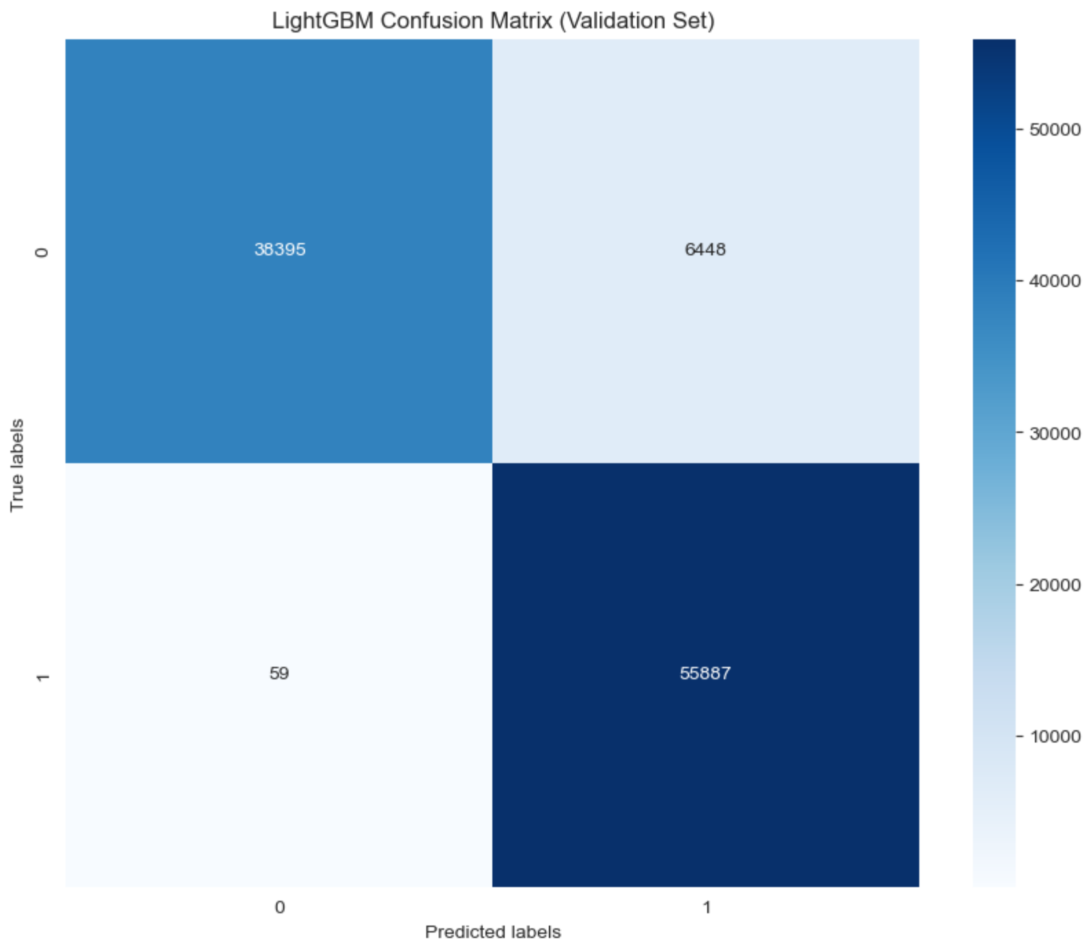

Figure 13: LightGBM Confusion Matrix on Validation Set

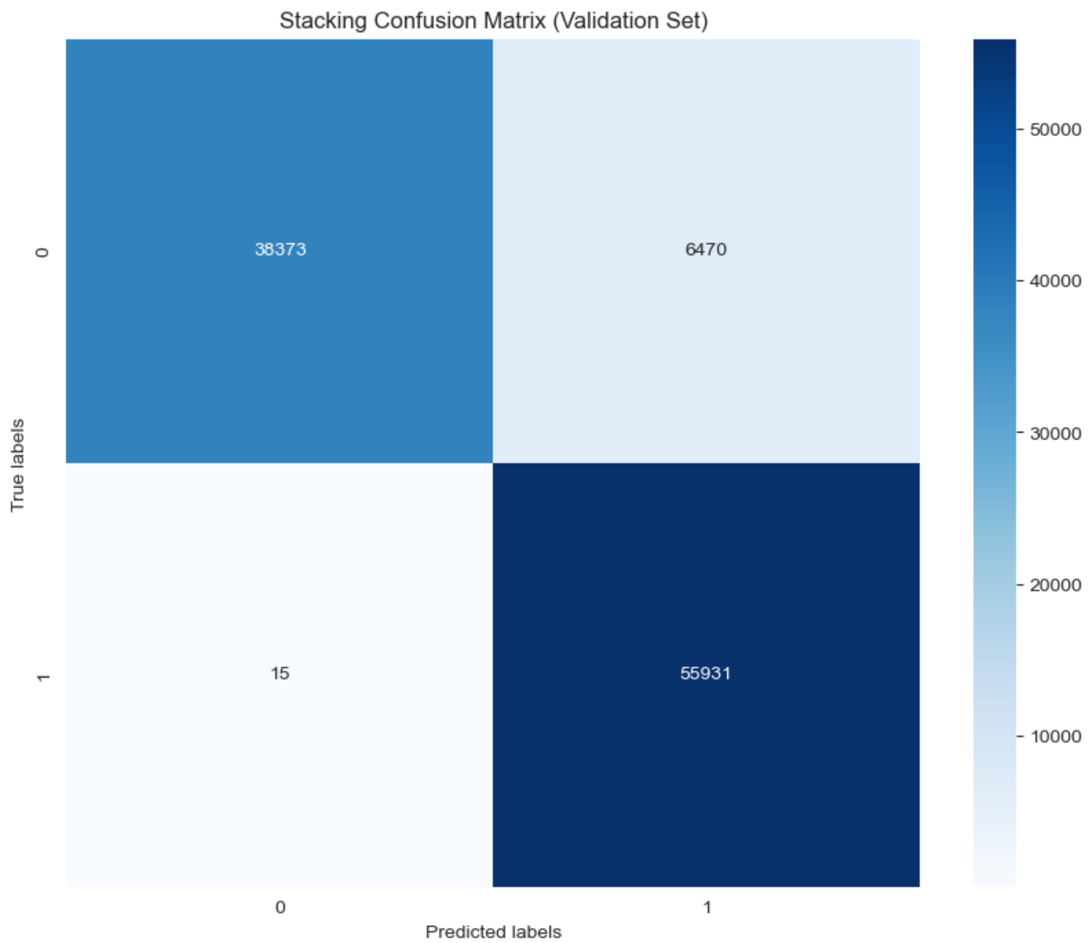

Figure 14: Stacked Confusion Matrix on Validation Set

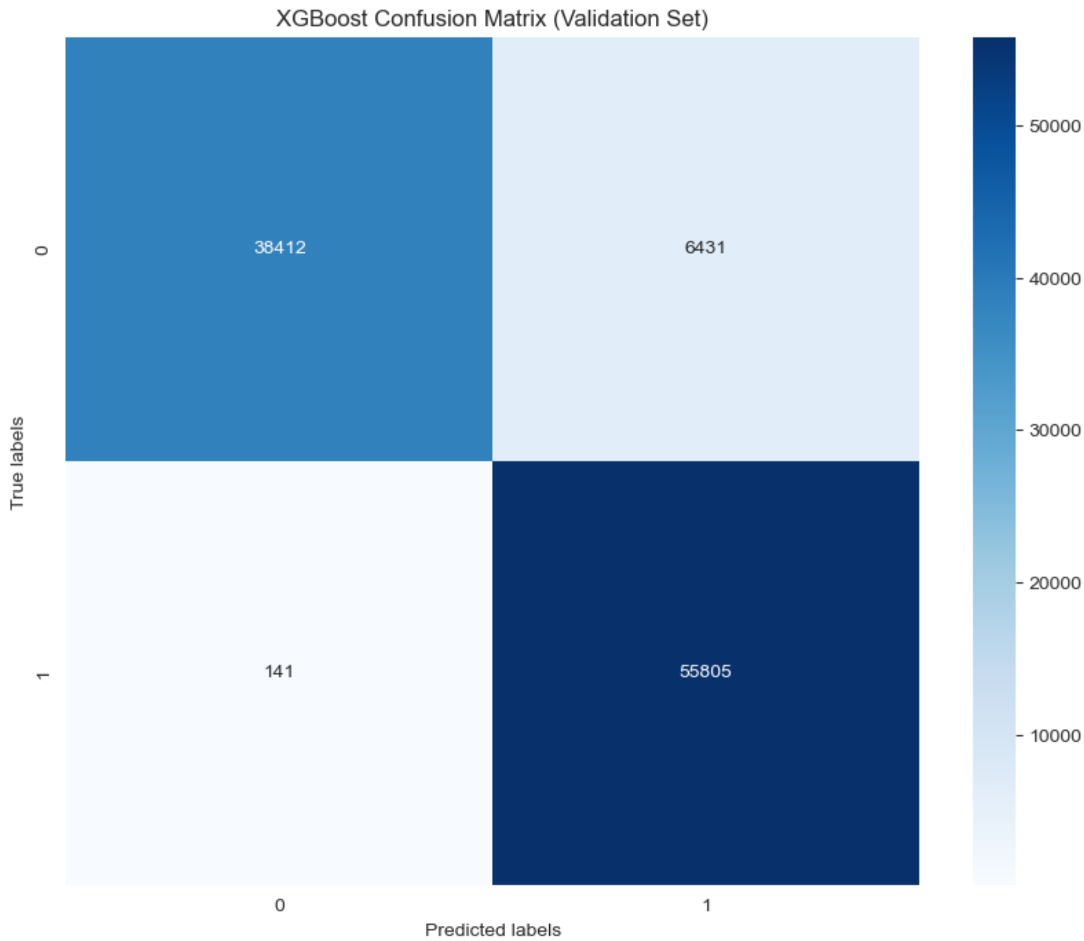

Figure 15: XGBoost Confusion Matrix on Validation Set

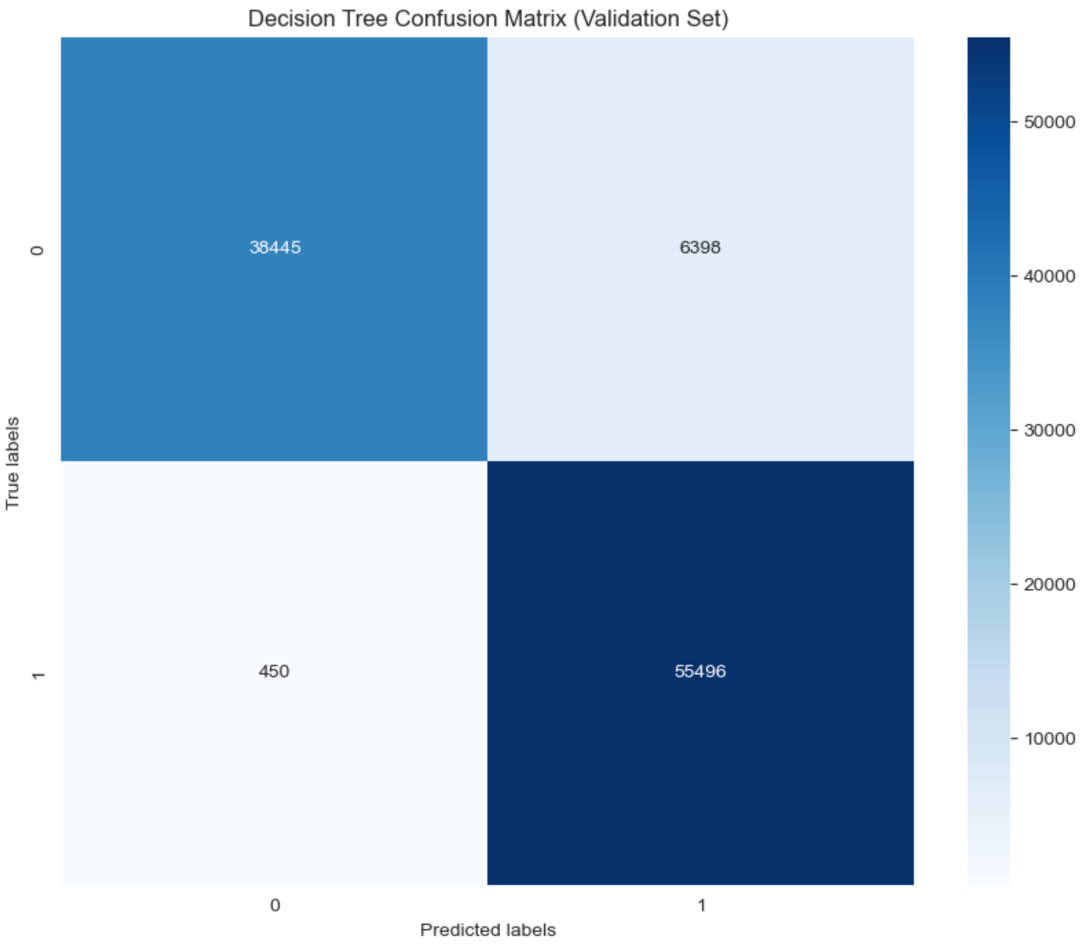

Figure 16: Decision Tree Confusion Matrix on Validation Set

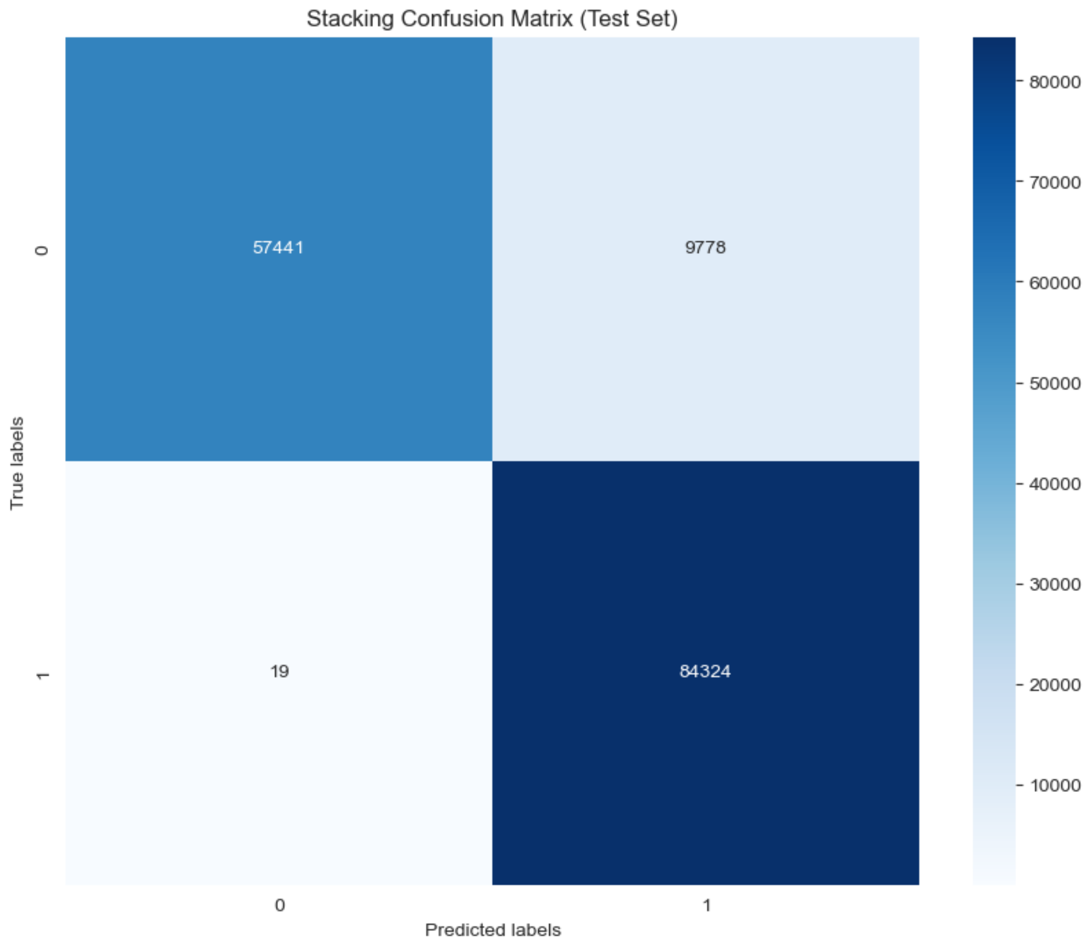

Figure 17: Stacked Confusion Matrix on Test Set

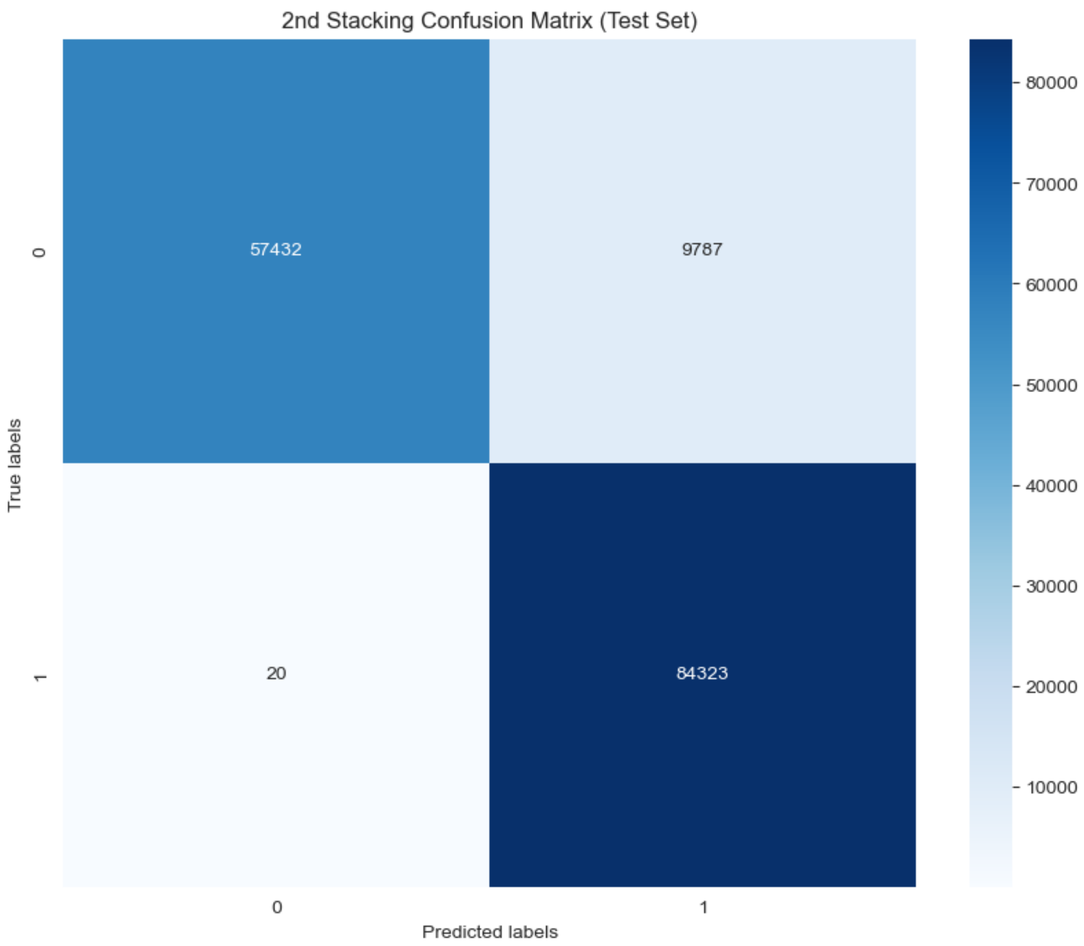

Figure 18: Hyperparameter tuned Stacked Confusion Matrix on Test Set

Age	Gender	Did not churn	Churned
(17.953, 33.667]	Female	38.025946%	61.974054%
(17.953, 33.667]	Male	55.880714%	44.119286%
(33.667, 49.333]	Female	46.674686%	53.325314%
(33.667, 49.333]	Male	64.828696%	35.171304%
(49.333, 65.0]	Female	13.456105%	86.543895%
(49.333, 65.0]	Male	19.450720%	80.549280%

Table 3: Cross-Tabulation for Age and Gender

Age	Tenure	Did not churn	Churned
(17.953, 33.667]	(0.941, 20.667]	46.438193%	53.561807%
(17.953, 33.667]	(20.667, 40.333]	48.112829%	51.887171%
(17.953, 33.667]	(40.333, 60.0]	49.270641%	50.729359%
(33.667, 49.333]	(0.941, 20.667]	55.417985%	44.582015%
(33.667, 49.333]	(20.667, 40.333]	57.593139%	42.406861%
(33.667, 49.333]	(40.333, 60.0]	58.264728%	41.735272%
(49.333, 65.0]	(0.941, 20.667]	17.309815%	82.690185%
(49.333, 65.0]	(20.667, 40.333]	16.208328%	83.791672%
(49.333, 65.0]	(40.333, 60.0]	16.010197%	83.989803%

Table 4: Cross-Tabulation for Age and Tenure

Age	Usage Frequency	Did not churn	Churned
(17.953, 33.667]	(0.971, 10.667]	44.344614%	55.655386%
(17.953, 33.667]	(10.667, 20.333]	49.631629%	50.368371%
(17.953, 33.667]	(20.333, 30.0]	49.752025%	50.247975%
(33.667, 49.333]	(0.971, 10.667]	52.827295%	47.172705%
(33.667, 49.333]	(10.667, 20.333]	59.358134%	40.641866%
(33.667, 49.333]	(20.333, 30.0]	58.919008%	41.080992%
(49.333, 65.0]	(0.971, 10.667]	13.234218%	86.765782%
(49.333, 65.0]	(10.667, 20.333]	18.058621%	81.941379%
(49.333, 65.0]	(20.333, 30.0]	18.240694%	81.759306%

Table 5: Cross-Tabulation for Age and Usage Frequency

Age	Support Calls	Did not churn	Churned
(17.953, 33.667]	(-0.01, 3.333]	72.974941%	27.025059%
(17.953, 33.667]	(3.333, 6.667]	10.175486%	89.824514%
(17.953, 33.667]	(6.667, 10.0]	9.083209%	90.916791%
(33.667, 49.333]	(-0.01, 3.333]	77.971783%	22.028217%
(33.667, 49.333]	(3.333, 6.667]	41.051369%	58.948631%
(33.667, 49.333]	(6.667, 10.0]	8.909835%	91.090165%
(49.333, 65.0]	(-0.01, 3.333]	25.048420%	74.951580%
(49.333, 65.0]	(3.333, 6.667]	16.305570%	83.694430%
(49.333, 65.0]	(6.667, 10.0]	7.497012%	92.502988%

Table 6: Cross-Tabulation for Age and Support Calls

Age	Payment Delay	Did not churn	Churned
(17.953, 33.667]	(-0.03, 10.0]	60.291204%	39.708796%
(17.953, 33.667]	(10.0, 20.0]	57.925140%	42.074860%
(17.953, 33.667]	(20.0, 30.0]	5.879965%	94.120035%
(33.667, 49.333]	(-0.03, 10.0]	69.071993%	30.928007%
(33.667, 49.333]	(10.0, 20.0]	67.289944%	32.710056%
(33.667, 49.333]	(20.0, 30.0]	5.898095%	94.101905%
(49.333, 65.0]	(-0.03, 10.0]	22.691953%	77.308047%
(49.333, 65.0]	(10.0, 20.0]	20.358381%	79.641619%
(49.333, 65.0]	(20.0, 30.0]	5.619901%	94.380099%

Table 7: Cross-Tabulation for Age and Payment Delay

Age	Subscription Type	Did not churn	Churned
(17.953, 33.667]	Basic	46.712369%	53.287631%
(17.953, 33.667]	Premium	48.633036%	51.366964%
(17.953, 33.667]	Standard	48.620684%	51.379316%
(33.667, 49.333]	Basic	55.726382%	44.273618%
(33.667, 49.333]	Premium	58.025383%	41.974617%
(33.667, 49.333]	Standard	57.676592%	42.323408%
(49.333, 65.0]	Basic	15.916008%	84.083992%
(49.333, 65.0]	Premium	16.810928%	83.189072%
(49.333, 65.0]	Standard	16.759833%	83.240167%

Table 8: Cross-Tabulation for Age and Subscription Type

Age	Contract Length	Did not churn	Churned
(17.953, 33.667]	Annual	57.414222%	42.585778%
(17.953, 33.667]	Monthly	10.011664%	89.988336%
(17.953, 33.667]	Quarterly	58.048079%	41.951921%
(33.667, 49.333]	Annual	66.867284%	33.132716%
(33.667, 49.333]	Monthly	10.275883%	89.724117%
(33.667, 49.333]	Quarterly	66.921615%	33.078385%
(49.333, 65.0]	Annual	19.741196%	80.258804%
(49.333, 65.0]	Monthly	9.139538%	90.860462%
(49.333, 65.0]	Quarterly	19.995036%	80.004964%

Table 9: Cross-Tabulation for Age and Contract Length

Age	Total Spend	Did not churn	Churned
(17.953, 33.667]	(99.1, 400.0]	10.155280%	89.844720%
(17.953, 33.667]	(400.0, 700.0]	52.326243%	47.673757%
(17.953, 33.667]	(700.0, 1000.0]	62.252162%	37.747838%
(33.667, 49.333]	(99.1, 400.0]	10.139011%	89.860989%
(33.667, 49.333]	(400.0, 700.0]	61.439981%	38.560019%
(33.667, 49.333]	(700.0, 1000.0]	71.304294%	28.695706%
(49.333, 65.0]	(99.1, 400.0]	9.263935%	90.736065%
(49.333, 65.0]	(400.0, 700.0]	17.880978%	82.119022%
(49.333, 65.0]	(700.0, 1000.0]	21.620901%	78.379099%

Table 10: Cross-Tabulation for Age and Total Spend

Age	Last Interaction	Did not churn	Churned
(17.953, 33.667]	(0.971, 10.667]	54.559529%	45.440471%
(17.953, 33.667]	(10.667, 20.333]	48.087580%	51.912420%
(17.953, 33.667]	(20.333, 30.0]	39.374315%	60.625685%
(33.667, 49.333]	(0.971, 10.667]	64.069158%	35.930842%
(33.667, 49.333]	(10.667, 20.333]	57.163841%	42.836159%
(33.667, 49.333]	(20.333, 30.0]	47.608710%	52.391290%
(49.333, 65.0]	(0.971, 10.667]	18.336108%	81.663892%
(49.333, 65.0]	(10.667, 20.333]	16.719798%	83.280202%
(49.333, 65.0]	(20.333, 30.0]	14.338245%	85.661755%

Table 11: Cross-Tabulation for Age and Last Interaction

Gender	Tenure	Did not churn	Churned
Female	(0.941, 20.667]	33.691029%	66.308971%
Female	(20.667, 40.333]	35.327635%	64.672365%
Female	(40.333, 60.0]	36.176015%	63.823985%
Male	(0.941, 20.667]	50.805936%	49.194064%
Male	(20.667, 40.333]	52.093927%	47.906073%
Male	(40.333, 60.0]	52.955252%	47.044748%

Table 12: Cross-Tabulation for Gender and Tenure

Gender	Usage Frequency	Did not churn	Churned
Female	(0.971, 10.667]	31.473896%	68.526104%
Female	(10.667, 20.333]	36.938864%	63.061136%
Female	(20.333, 30.0]	36.762487%	63.237513%
Male	(0.971, 10.667]	47.355478%	52.644522%
Male	(10.667, 20.333]	54.164282%	45.835718%
Male	(20.333, 30.0]	54.112338%	45.887662%

Table 13: Cross-Tabulation for Gender and Usage Frequency

Gender	**Support Calls**	**Did not churn**	**Churned**
Female	(-0.01, 3.333]	56.847982%	43.152018%
Female	(3.333, 6.667]	19.473340%	80.526660%
Female	(6.667, 10.0]	7.947850%	92.052150%
Male	(-0.01, 3.333]	75.122383%	24.877617%
Male	(3.333, 6.667]	29.579506%	70.420494%
Male	(6.667, 10.0]	9.045318%	90.954682%

Table 14: Cross-Tabulation for Gender and Support Calls

Gender	**Payment Delay**	**Did not churn**	**Churned**
Female	(-0.03, 10.0]	47.567799%	52.432201%
Female	(10.0, 20.0]	44.950655%	55.049345%
Female	(20.0, 30.0]	2.916100%	97.083900%
Male	(-0.03, 10.0]	63.200635%	36.799365%
Male	(10.0, 20.0]	61.136121%	38.863879%
Male	(20.0, 30.0]	8.918586%	91.081414%

Table 15: Cross-Tabulation for Gender and Payment Delay

Gender	**Subscription Type**	**Did not churn**	**Churned**
Female	Basic	33.951525%	66.048475%
Female	Premium	35.642264%	64.357736%
Female	Standard	35.687542%	64.312458%
Male	Basic	50.503641%	49.496359%
Male	Premium	52.891000%	47.109000%
Male	Standard	52.556845%	47.443155%

Table 16: Cross-Tabulation for Gender and Subscription Type

Gender	**Contract Length**	**Did not churn**	**Churned**
Female	Annual	43.850312%	56.149688%
Female	Monthly	8.256848%	91.743152%
Female	Quarterly	44.069740%	55.930260%
Male	Annual	61.364732%	38.635268%
Male	Monthly	11.418148%	88.581852%
Male	Quarterly	61.770601%	38.229399%

Table 17: Cross-Tabulation for Gender and Contract Length

Gender	Total Spend	Did not churn	Churned
Female	(99.1, 400.0]	8.206611%	91.793389%
Female	(400.0, 700.0]	38.843479%	61.156521%
Female	(700.0, 1000.0]	48.414524%	51.585476%
Male	(99.1, 400.0]	11.555597%	88.444403%
Male	(400.0, 700.0]	56.255421%	43.744579%
Male	(700.0, 1000.0]	65.889919%	34.110081%

Table 18: Cross-Tabulation for Gender and Total Spend

Gender	Last Interaction	Did not churn	Churned
Female	(0.971, 10.667]	50.002105%	49.997895%
Female	(10.667, 20.333]	35.121364%	64.878636%
Female	(20.333, 30.0]	9.412746%	90.587254%
Male	(0.971, 10.667]	51.904262%	48.095738%
Male	(10.667, 20.333]	52.106489%	47.893511%
Male	(20.333, 30.0]	51.997684%	48.002316%

Table 19: Cross-Tabulation for Gender and Last Interaction

Tenure	Usage Frequency	Did not churn	Churned
(0.941, 20.667]	(0.971, 10.667]	33.432848%	66.567152%
(0.941, 20.667]	(10.667, 20.333]	47.379933%	52.620067%
(0.941, 20.667]	(20.333, 30.0]	47.042664%	52.957336%
(20.667, 40.333]	(0.971, 10.667]	40.678462%	59.321538%
(20.667, 40.333]	(10.667, 20.333]	46.599622%	53.400378%
(20.667, 40.333]	(20.333, 30.0]	46.467288%	53.532712%
(40.333, 60.0]	(0.971, 10.667]	45.054389%	54.945611%
(40.333, 60.0]	(10.667, 20.333]	45.793136%	54.206864%
(40.333, 60.0]	(20.333, 30.0]	45.827057%	54.172943%

Table 20: Cross-Tabulation for Tenure and Usage Frequency

Tenure	Support Calls	Did not churn	Churned
(0.941, 20.667]	(-0.01, 3.333]	64.201117%	35.798883%
(0.941, 20.667]	(3.333, 6.667]	25.641102%	74.358898%
(0.941, 20.667]	(6.667, 10.0]	11.977766%	88.022234%
(20.667, 40.333]	(-0.01, 3.333]	68.424278%	31.575722%
(20.667, 40.333]	(3.333, 6.667]	24.320894%	75.679106%
(20.667, 40.333]	(6.667, 10.0]	7.282954%	92.717046%
(40.333, 60.0]	(-0.01, 3.333]	69.991898%	30.008102%
(40.333, 60.0]	(3.333, 6.667]	24.351778%	75.648222%
(40.333, 60.0]	(6.667, 10.0]	6.448147%	93.551853%

Table 21: Cross-Tabulation for Tenure and Support Calls

Tenure	Payment Delay	Did not churn	Churned
(0.941, 20.667]	(-0.03, 10.0]	53.098302%	46.901698%
(0.941, 20.667]	(10.0, 20.0]	53.082684%	46.917316%
(0.941, 20.667]	(20.0, 30.0]	9.013272%	90.986728%
(20.667, 40.333]	(-0.03, 10.0]	57.470799%	42.529201%
(20.667, 40.333]	(10.0, 20.0]	54.250338%	45.749662%
(20.667, 40.333]	(20.0, 30.0]	4.620181%	95.379819%
(40.333, 60.0]	(-0.03, 10.0]	58.745588%	41.254412%
(40.333, 60.0]	(10.0, 20.0]	55.338753%	44.661247%
(40.333, 60.0]	(20.0, 30.0]	3.964746%	96.035254%

Table 22: Cross-Tabulation for Tenure and Payment Delay

Tenure	Subscription Type	Did not churn	Churned
(0.941, 20.667]	Basic	38.107412%	61.892588%
(0.941, 20.667]	Premium	45.412404%	54.587596%
(0.941, 20.667]	Standard	45.207990%	54.792010%
(20.667, 40.333]	Basic	44.872494%	55.127506%
(20.667, 40.333]	Premium	44.543145%	55.456855%
(20.667, 40.333]	Standard	44.565674%	55.434326%
(40.333, 60.0]	Basic	45.576923%	54.423077%
(40.333, 60.0]	Premium	45.779408%	54.220592%
(40.333, 60.0]	Standard	45.317830%	54.682170%

Table 23: Cross-Tabulation for Tenure and Subscription Type

Tenure	Contract Length	Did not churn	Churned
(0.941, 20.667]	Annual	52.314724%	47.685276%
(0.941, 20.667]	Monthly	11.252995%	88.747005%
(0.941, 20.667]	Quarterly	52.386852%	47.613148%
(20.667, 40.333]	Annual	54.021962%	45.978038%
(20.667, 40.333]	Monthly	9.387900%	90.612100%
(20.667, 40.333]	Quarterly	54.495544%	45.504456%
(40.333, 60.0]	Annual	55.183490%	44.816510%
(40.333, 60.0]	Monthly	8.830016%	91.169984%
(40.333, 60.0]	Quarterly	55.443667%	44.556333%

Table 24: Cross-Tabulation for Tenure and Contract Length

Tenure	**Total Spend**	**Did not churn**	**Churned**
(0.941, 20.667]	(99.1, 400.0]	11.360824%	88.639176%
(0.941, 20.667]	(400.0, 700.0]	46.925013%	53.074987%
(0.941, 20.667]	(700.0, 1000.0]	56.919131%	43.080869%
(20.667, 40.333]	(99.1, 400.0]	9.361946%	90.638054%
(20.667, 40.333]	(400.0, 700.0]	48.954215%	51.045785%
(20.667, 40.333]	(700.0, 1000.0]	58.705898%	41.294102%
(40.333, 60.0]	(99.1, 400.0]	8.884703%	91.115297%
(40.333, 60.0]	(400.0, 700.0]	49.848430%	50.151570%
(40.333, 60.0]	(700.0, 1000.0]	59.835690%	40.164310%

Table 25: Cross-Tabulation for Tenure and Total Spend

Tenure	**Last Interaction**	**Did not churn**	**Churned**
(0.941, 20.667]	(0.971, 10.667]	49.230954%	50.769046%
(0.941, 20.667]	(10.667, 20.333]	43.093291%	56.906709%
(0.941, 20.667]	(20.333, 30.0]	35.313759%	64.686241%
(20.667, 40.333]	(0.971, 10.667]	51.227406%	48.772594%
(20.667, 40.333]	(10.667, 20.333]	44.781790%	55.218210%
(20.667, 40.333]	(20.333, 30.0]	36.153892%	63.846108%
(40.333, 60.0]	(0.971, 10.667]	52.177323%	47.822677%
(40.333, 60.0]	(10.667, 20.333]	45.655009%	54.344991%
(40.333, 60.0]	(20.333, 30.0]	36.974271%	63.025729%

Table 26: Cross-Tabulation for Tenure and Last Interaction

Usage Frequency	**Support Calls**	**Did not churn**	**Churned**
(0.971, 10.667]	(-0.01, 3.333]	63.838457%	36.161543%
(0.971, 10.667]	(3.333, 6.667]	21.210204%	78.789796%
(0.971, 10.667]	(6.667, 10.0]	7.322505%	92.677495%
(10.667, 20.333]	(-0.01, 3.333]	69.393230%	30.606770%
(10.667, 20.333]	(3.333, 6.667]	26.557767%	73.442233%
(10.667, 20.333]	(6.667, 10.0]	9.113442%	90.886558%
(20.333, 30.0]	(-0.01, 3.333]	69.400523%	30.599477%
(20.333, 30.0]	(3.333, 6.667]	26.422467%	73.577533%
(20.333, 30.0]	(6.667, 10.0]	9.070785%	90.929215%

Table 27: Cross-Tabulation for Usage Frequency and Support Calls

Usage Frequency	Payment Delay	Did not churn	Churned
(0.971, 10.667]	(-0.03, 10.0]	51.743168%	48.256832%
(0.971, 10.667]	(10.0, 20.0]	49.653259%	50.346741%
(0.971, 10.667]	(20.0, 30.0]	5.209864%	94.790136%
(10.667, 20.333]	(-0.03, 10.0]	58.841025%	41.158975%
(10.667, 20.333]	(10.0, 20.0]	56.533292%	43.466708%
(10.667, 20.333]	(20.0, 30.0]	6.087800%	93.912200%
(20.333, 30.0]	(-0.03, 10.0]	58.718459%	41.281541%
(20.333, 30.0]	(10.0, 20.0]	56.268428%	43.731572%
(20.333, 30.0]	(20.0, 30.0]	6.103893%	93.896107%

Table 28: Cross-Tabulation for Usage Frequency and Payment Delay

Usage Frequency	Subscription Type	Did not churn	Churned
(0.971, 10.667]	Basic	38.576380%	61.423620%
(0.971, 10.667]	Premium	40.993193%	59.006807%
(0.971, 10.667]	Standard	40.872819%	59.127181%
(10.667, 20.333]	Basic	45.558642%	54.441358%
(10.667, 20.333]	Premium	47.181740%	52.818260%
(10.667, 20.333]	Standard	46.939762%	53.060238%
(20.333, 30.0]	Basic	44.925033%	55.074967%
(20.333, 30.0]	Premium	47.294260%	52.705740%
(20.333, 30.0]	Standard	47.027371%	52.972629%

Table 29: Cross-Tabulation for Usage Frequency and Subscription Type

Usage Frequency	Contract Length	Did not churn	Churned
(0.971, 10.667]	Annual	49.111008%	50.888992%
(0.971, 10.667]	Monthly	8.917301%	91.082699%
(0.971, 10.667]	Quarterly	49.679202%	50.320798%
(10.667, 20.333]	Annual	56.095909%	43.904091%
(10.667, 20.333]	Monthly	10.297331%	89.702669%
(10.667, 20.333]	Quarterly	56.386983%	43.613017%
(20.333, 30.0]	Annual	56.137775%	43.862225%
(20.333, 30.0]	Monthly	10.209532%	89.790468%
(20.333, 30.0]	Quarterly	56.093132%	43.906868%

Table 30: Cross-Tabulation for Usage Frequency and Contract Length

Usage Frequency	**Total Spend**	**Did not churn**	**Churned**
(0.971, 10.667]	(99.1, 400.0]	8.675079%	91.324921%
(0.971, 10.667]	(400.0, 700.0]	44.283349%	55.716651%
(0.971, 10.667]	(700.0, 1000.0]	53.905097%	46.094903%
(10.667, 20.333]	(99.1, 400.0]	10.406428%	89.593572%
(10.667, 20.333]	(400.0, 700.0]	50.867711%	49.132289%
(10.667, 20.333]	(700.0, 1000.0]	60.677687%	39.322313%
(20.333, 30.0]	(99.1, 400.0]	10.483028%	89.516972%
(20.333, 30.0]	(400.0, 700.0]	50.494098%	49.505902%
(20.333, 30.0]	(700.0, 1000.0]	60.645578%	39.354422%

Table 31: Cross-Tabulation for Usage Frequency and Total Spend

Usage Frequency	**Last Interaction**	**Did not churn**	**Churned**
(0.971, 10.667]	(0.971, 10.667]	46.617439%	53.382561%
(0.971, 10.667]	(10.667, 20.333]	40.104893%	59.895107%
(0.971, 10.667]	(20.333, 30.0]	32.244090%	67.755910%
(10.667, 20.333]	(0.971, 10.667]	53.065976%	46.934024%
(10.667, 20.333]	(10.667, 20.333]	46.544910%	53.455090%
(10.667, 20.333]	(20.333, 30.0]	38.227402%	61.772598%
(20.333, 30.0]	(0.971, 10.667]	52.813792%	47.186208%
(20.333, 30.0]	(10.667, 20.333]	46.763043%	53.236957%
(20.333, 30.0]	(20.333, 30.0]	37.891337%	62.108663%

Table 32: Cross-Tabulation for Usage Frequency and Last Interaction

Support Calls	**Payment Delay**	**Did not churn**	**Churned**
(-0.01, 3.333]	(-0.03, 10.0]	77.083317%	22.916683%
(-0.01, 3.333]	(10.0, 20.0]	77.040928%	22.959072%
(-0.01, 3.333]	(20.0, 30.0]	9.417844%	90.582156%
(3.333, 6.667]	(-0.03, 10.0]	34.319850%	65.680150%
(3.333, 6.667]	(10.0, 20.0]	31.365445%	68.634555%
(3.333, 6.667]	(20.0, 30.0]	5.004381%	94.995619%
(6.667, 10.0]	(-0.03, 10.0]	13.596909%	86.403091%
(6.667, 10.0]	(10.0, 20.0]	9.137509%	90.862491%
(6.667, 10.0]	(20.0, 30.0]	3.088430%	96.911570%

Table 33: Cross-Tabulation for Support Calls and Payment Delay

Support Calls	Subscription Type	Did not churn	Churned
(-0.01, 3.333]	Basic	66.279161%	33.720839%
(-0.01, 3.333]	Premium	68.419936%	31.580064%
(-0.01, 3.333]	Standard	68.306964%	31.693036%
(3.333, 6.667]	Basic	24.051106%	75.948894%
(3.333, 6.667]	Premium	25.392722%	74.607278%
(3.333, 6.667]	Standard	24.797622%	75.202378%
(6.667, 10.0]	Basic	8.464329%	91.535671%
(6.667, 10.0]	Premium	8.536222%	91.463778%
(6.667, 10.0]	Standard	8.470368%	91.529632%

Table 34: Cross-Tabulation for Support Calls and Subscription Type

Support Calls	Contract Length	Did not churn	Churned
(-0.01, 3.333]	Annual	76.579358%	23.420642%
(-0.01, 3.333]	Monthly	12.025821%	87.974179%
(-0.01, 3.333]	Quarterly	76.909414%	23.090586%
(3.333, 6.667]	Annual	30.916251%	69.083749%
(3.333, 6.667]	Monthly	9.294355%	90.705645%
(3.333, 6.667]	Quarterly	30.709224%	69.290776%
(6.667, 10.0]	Annual	8.644130%	91.355870%
(6.667, 10.0]	Monthly	8.048582%	91.951418%
(6.667, 10.0]	Quarterly	8.780354%	91.219646%

Table 35: Cross-Tabulation for Support Calls and Contract Length

Support Calls	Total Spend	Did not churn	Churned
(-0.01, 3.333]	(99.1, 400.0]	12.021491%	87.978509%
(-0.01, 3.333]	(400.0, 700.0]	71.810832%	28.189168%
(-0.01, 3.333]	(700.0, 1000.0]	80.440603%	19.559397%
(3.333, 6.667]	(99.1, 400.0]	9.317577%	90.682423%
(3.333, 6.667]	(400.0, 700.0]	26.983200%	73.016800%
(3.333, 6.667]	(700.0, 1000.0]	34.418007%	65.581993%
(6.667, 10.0]	(99.1, 400.0]	8.171964%	91.828036%
(6.667, 10.0]	(400.0, 700.0]	8.510481%	91.489519%
(6.667, 10.0]	(700.0, 1000.0]	8.789793%	91.210207%

Table 36: Cross-Tabulation for Support Calls and Total Spend

Support Calls	**Last Interaction**	**Did not churn**	**Churned**
(-0.01, 3.333]	(0.971, 10.667]	73.974049%	26.025951%
(-0.01, 3.333]	(10.667, 20.333]	67.636791%	32.363209%
(-0.01, 3.333]	(20.333, 30.0]	58.420190%	41.579810%
(3.333, 6.667]	(0.971, 10.667]	28.774706%	71.225294%
(3.333, 6.667]	(10.667, 20.333]	24.702980%	75.297020%
(3.333, 6.667]	(20.333, 30.0]	20.339897%	79.660103%
(6.667, 10.0]	(0.971, 10.667]	8.289389%	91.710611%
(6.667, 10.0]	(10.667, 20.333]	8.775764%	91.224236%
(6.667, 10.0]	(20.333, 30.0]	8.407624%	91.592376%

Table 37: Cross-Tabulation for Support Calls and Last Interaction

Payment Delay	**Subscription Type**	**Did not churn**	**Churned**
(-0.03, 10.0]	Basic	55.176871%	44.823129%
(-0.03, 10.0]	Premium	57.177485%	42.822515%
(-0.03, 10.0]	Standard	57.284927%	42.715073%
(10.0, 20.0]	Basic	52.732081%	47.267919%
(10.0, 20.0]	Premium	55.208236%	44.791764%
(10.0, 20.0]	Standard	54.827425%	45.172575%
(20.0, 30.0]	Basic	5.735473%	94.264527%
(20.0, 30.0]	Premium	5.874424%	94.125576%
(20.0, 30.0]	Standard	5.787216%	94.212784%

Table 38: Cross-Tabulation for Payment Delay and Subscription Type

Payment Delay	**Contract Length**	**Did not churn**	**Churned**
(-0.03, 10.0]	Annual	65.624627%	34.375373%
(-0.03, 10.0]	Monthly	14.231234%	85.768766%
(-0.03, 10.0]	Quarterly	65.992530%	34.007470%
(10.0, 20.0]	Annual	63.580785%	36.419215%
(10.0, 20.0]	Monthly	11.486721%	88.513279%
(10.0, 20.0]	Quarterly	64.046715%	35.953285%
(20.0, 30.0]	Annual	6.842591%	93.157409%
(20.0, 30.0]	Monthly	4.044985%	95.955015%
(20.0, 30.0]	Quarterly	6.573819%	93.426181%

Table 39: Cross-Tabulation for Payment Delay and Contract Length

Payment Delay	Total Spend	Did not churn	Churned
(-0.03, 10.0]	(99.1, 400.0]	14.019997%	85.980003%
(-0.03, 10.0]	(400.0, 700.0]	60.557505%	39.442495%
(-0.03, 10.0]	(700.0, 1000.0]	70.112231%	29.887769%
(10.0, 20.0]	(99.1, 400.0]	11.483886%	88.516114%
(10.0, 20.0]	(400.0, 700.0]	58.823103%	41.176897%
(10.0, 20.0]	(700.0, 1000.0]	67.908804%	32.091196%
(20.0, 30.0]	(99.1, 400.0]	4.349770%	95.650230%
(20.0, 30.0]	(400.0, 700.0]	6.019689%	93.980311%
(20.0, 30.0]	(700.0, 1000.0]	7.083767%	92.916233%

Table 40: Cross-Tabulation for Payment Delay and Total Spend

Payment Delay	Last Interaction	Did not churn	Churned
(-0.03, 10.0]	(0.971, 10.667]	63.112499%	36.887501%
(-0.03, 10.0]	(10.667, 20.333]	56.604913%	43.395087%
(-0.03, 10.0]	(20.333, 30.0]	47.609006%	52.390994%
(10.0, 20.0]	(0.971, 10.667]	61.111340%	38.888660%
(10.0, 20.0]	(10.667, 20.333]	54.274929%	45.725071%
(10.0, 20.0]	(20.333, 30.0]	45.025439%	54.974561%
(20.0, 30.0]	(0.971, 10.667]	5.701145%	94.298855%
(20.0, 30.0]	(10.667, 20.333]	5.973928%	94.026072%
(20.0, 30.0]	(20.333, 30.0]	5.722738%	94.277262%

Table 41: Cross-Tabulation for Payment Delay and Last Interaction

Subscription Type	Contract Length	Did not churn	Churned
Basic	Annual	52.585151%	47.414849%
Basic	Monthly	9.694424%	90.305576%
Basic	Quarterly	52.652201%	47.347799%
Premium	Annual	54.582007%	45.417993%
Premium	Monthly	9.978617%	90.021383%
Premium	Quarterly	55.171018%	44.828982%
Standard	Annual	54.508794%	45.491206%
Standard	Monthly	9.737704%	90.262296%
Standard	Quarterly	54.650868%	45.349132%

Table 42: Cross-Tabulation for Subscription Type and Contract Length

Subscription Type	**Total Spend**	**Did not churn**	**Churned**
Basic	(99.1, 400.0]	9.607239%	90.392761%
Basic	(400.0, 700.0]	47.171678%	52.828322%
Basic	(700.0, 1000.0]	57.220783%	42.779217%
Premium	(99.1, 400.0]	10.003590%	89.996410%
Premium	(400.0, 700.0]	49.314844%	50.685156%
Premium	(700.0, 1000.0]	59.341804%	40.658196%
Standard	(99.1, 400.0]	9.937736%	90.062264%
Standard	(400.0, 700.0]	49.393360%	50.606640%
Standard	(700.0, 1000.0]	59.051224%	40.948776%

Table 43: Cross-Tabulation for Subscription Type and Total Spend

Subscription Type	**Last Interaction**	**Did not churn**	**Churned**
Basic	(0.971, 10.667]	49.591049%	50.408951%
Basic	(10.667, 20.333]	43.119049%	56.880951%
Basic	(20.333, 30.0]	34.927093%	65.072907%
Premium	(0.971, 10.667]	51.606404%	48.393596%
Premium	(10.667, 20.333]	45.420620%	54.579380%
Premium	(20.333, 30.0]	36.933063%	63.066937%
Standard	(0.971, 10.667]	51.584278%	48.415722%
Standard	(10.667, 20.333]	45.100429%	54.899571%
Standard	(20.333, 30.0]	36.634156%	63.365844%

Table 44: Cross-Tabulation for Subscription Type and Last Interaction

Contract Length	**Total Spend**	**Did not churn**	**Churned**
Annual	(99.1, 400.0]	9.708208%	90.291792%
Annual	(400.0, 700.0]	58.154102%	41.845898%
Annual	(700.0, 1000.0]	68.550202%	31.449798%
Monthly	(99.1, 400.0]	10.040050%	89.959950%
Monthly	(400.0, 700.0]	9.896445%	90.103555%
Monthly	(700.0, 1000.0]	9.475707%	90.524293%
Quarterly	(99.1, 400.0]	9.802846%	90.197154%
Quarterly	(400.0, 700.0]	58.883277%	41.116723%
Quarterly	(700.0, 1000.0]	68.526471%	31.473529%

Table 45: Cross-Tabulation for Contract Length and Total Spend

Contract Length	**Last Interaction**	**Did not churn**	**Churned**
Annual	(0.971, 10.667]	60.793385%	39.206615%
Annual	(10.667, 20.333]	54.101367%	45.898633%
Annual	(20.333, 30.0]	44.355870%	55.644130%
Monthly	(0.971, 10.667]	9.635645%	90.364355%
Monthly	(10.667, 20.333]	10.050720%	89.949280%
Monthly	(20.333, 30.0]	9.724326%	90.275674%
Quarterly	(0.971, 10.667]	60.909259%	39.090741%
Quarterly	(10.667, 20.333]	54.114433%	45.885567%
Quarterly	(20.333, 30.0]	45.138565%	54.861435%

Table 46: Cross-Tabulation for Contract Length and Last Interaction

Total Spend	**Last Interaction**	**Did not churn**	**Churned**
(99.1, 400.0]	(0.971, 10.667]	9.501754%	90.498246%
(99.1, 400.0]	(10.667, 20.333]	9.988459%	90.011541%
(99.1, 400.0]	(20.333, 30.0]	10.059237%	89.940763%
(400.0, 700.0]	(0.971, 10.667]	55.258289%	44.741711%
(400.0, 700.0]	(10.667, 20.333]	48.753209%	51.246791%
(400.0, 700.0]	(20.333, 30.0]	39.879171%	60.120829%
(700.0, 1000.0]	(0.971, 10.667]	65.476077%	34.523923%
(700.0, 1000.0]	(10.667, 20.333]	58.553025%	41.446975%
(700.0, 1000.0]	(20.333, 30.0]	48.872394%	51.127606%

Table 47: Cross-Tabulation for Total Spend and Last Interaction

Metric	**Score**
Recall	0.8203920729205687
Precision	0.8822728231886282
F1 Score	0.8502079674646456
ROC AUC	0.841978344545920
Accuracy	0.8396112507670078

Table 48: Performance of baseline model

Metric	**Validation Set**	**Test Set**
Recall	0.9972473456547385	0.997166332712851
Precision	0.8967035792924991	0.8963444527336673
F1 Score	0.9443066897981636	0.9440712116875171
ROC AUC	0.9269625440001273	0.926237549789681
Accuracy	0.9347051761600969	0.934251329488922

Table 49: Performance of Random Forest on Validation Set and Test Set

Metric	Validation Set	Test Set
Recall	0.9974797125799878	0.997522023167305
Precision	0.8966675236197699	0.8963011889035667
F1 Score	0.9443908547833002	0.9442065865743415
ROC AUC	0.9270341274136921	0.9263558880322756
Accuracy	0.9347944716189267	0.9343964846069596

Table 50: Performance of XGBoost on Validation Set and Test Set

Metric	Validation Set	Test Set
Recall	0.9989454116469453	0.998849934197266
Precision	0.8965589155370177	0.8962817171126124
F1 Score	0.9449869378852056	0.9447902297883795
ROC AUC	0.9275774267386657	0.9269082679510706
Accuracy	0.935439383266031	0.9350364867183067

Table 51: Performance of LightGBM on Validation Set and Test Set

Metric	Validation Set	Test Set
Recall	0.9918671576162729	0.9912500148204356
Precision	0.8966793245536075	0.8960302660064733
F1 Score	0.9418743794077961	0.9412380593192194
ROC AUC	0.9246404003856402	0.9234653501704493
Accuracy	0.9320461558304973	0.931123896491205

Table 52: Performance of Decision Tree on Validation Set and Test Set

Metric	Validation Set	Test Set
Recall	0.99973188431702	0.9997747293788459
Precision	0.896315764170446	0.8960914752077533
F1 Score	0.9452035117070986	0.9450979293339685
ROC AUC	0.9277253628038726	0.9271549527225684
Accuracy	0.9356576610542817	0.9353597867539357

Table 53: Performance of Stacking Model on Validation and Test Set

Metric	Validation Set	Test Set
Recall	0.9996961355592893	0.9997628730303642
Precision	0.8962693503413353	0.8960046753798746
F1 Score	0.9451617264339067	0.945044353415185
ROC AUC	0.9276740383882123	0.927082079190616
Accuracy	0.935608052466043	0.9352938071548277

Table 54: Performance of Hyperparameter tuned Stacking Model on Validation and Test Set

References

Behl, A. R., Hadpe, H., Bonde, K., & Patel, H. (2024, Apr.). Customer churn prediction for a telecommunication company. *International Journal of Research in Engineering, Science and Management*, *7*(3), 107–113. Retrieved from https://journal.ijresm.com/index.php/ijresm/article/view/2973 doi: 10.5281/zenodo.10927526

Bhadane, P. S. N., Randhir, G., Borade, M., Bhatia, S., & More, G. (2023). Dynamic churn prediction system using machine learning algorithms. *International Research Journal of Modernization in Engineering Technology and Science*, *5*. Retrieved from https://www.irjmets.com/uploadedfiles/paper//issue_5_may_2023/40386/final/fin_irjmets1685177911.pdf

Dai, X. (2023, Oct.). Machine learning based churn prediction in telecom. *Highlights in Science, Engineering and Technology*, *68*, 212–222. Retrieved from https://drpress.org/ojs/index.php/HSET/article/view/12068 doi: 10.54097/hset.v68i.12068

Elechi, P., & Michael, O.-A. (2023, 07). Enhanced predictive data mining algorithm for fraud detection and churn behaviour modelling in telecommunication systems. *International Journal of Advanced Engineering and Technology*, *7*, 1-4. Retrieved from https://www.researchgate.net/publication/373734998_Enhanced_predictive_data_mining_algorithm_for_fraud_detection_and_churn_behaviour_modelling_in_telecommunication_systems

Oludele, A., Ben, A., A.C., O., S.O., K., & Seun, E. (2020). Enhanced churn prediction in the telecommunication industry. *International Journal of Innovative Research in Computer Science & Technology*. Retrieved from https://www.ijircst.org/view_abstract.php?title=Enhanced-Churn-Prediction-in-the-Telecommunication-Industry&year=2020&vol=8&primary=QVJULTM3Nw==#nav-cites doi: https://doi.org/10.21276/ijircst.2020.8.2.1

Summaila, & Stephen. (2023). Performance analysis of qos parameters and churn prediction model development for mnos in nigeria. *Journal of Systematic, Evaluation and Diversity Engineering*, *2*(2). Retrieved from https://africanscholarpub.com/ajsede/article/view/74

Usman-Hamza, A. O. B. S. K. N. L. F. C. H. A. M. S. A. S. A. G. A. M. A. M. J. B. A. F. E. (2024). Empirical analysis of tree-based classification models for customer churn prediction. *Scientific African*. Retrieved from https://www.sciencedirect.com/science/article/pii/S2468227623005070 doi: http://doi.org/10.1016/j.sciaf.2023.e02054

Zatonatska, T., Fareniuk, Y., & Shpyrko, V. (2023). Churn rate modeling for telecommunication operators using data science methods. *Marketing and Management of Innovations*, *14*(2), 163-173. Retrieved from https://mmi.sumdu.edu.ua/wp-content/uploads/2023/06/A690-2023_15_Zatonatska-et-al-3.pdf