School of Computing, Engineering and Digital Technologies
Department of Computing and Games
Teesside University,
Middlesbrough, TS1 3BA

Enhancing Explainability in Machine Learning Models for Customer Churn Prediction

Student Name: Lukka Sai Kumari

Student ID: S3308548

Course: MSc Computer Science (with Advanced Practice)

Supervisor: Han, The Anh

Date of submission: 06 Jan 2026

Declaration

I hereby declare that the work presented in this report, titled *"Enhancing Explainability in Machine Learning Models for Customer Churn Prediction,"* is my own and has not been submitted, in whole or in part, for any other academic award or qualification at Teesside University or any other institution.

All sources of information and references used in this project have been duly acknowledged in accordance with the university's academic integrity and referencing guidelines.

I confirm that this research complies with Teesside University's ethical standards and that the dataset used the Maven Telecom Customer Churn dataset is publicly available, anonymised, and suitable for academic research.

Signed By:

Sai kumari Lukka,

06 Jan 2026

Acknowledgement

I would like to extend my heartfelt thanks to my supervisor**, Mr. The Anh Han**, for his unwavering support, encouragement, and valuable feedback throughout this research project. His expertise in machine learning, explainable AI, and research methodology has been instrumental in shaping both the direction and the quality of this work.

I would also like to thank the **Department of Computing and Games, School of Computing, Engineering and Digital Technologies at Teesside University**, for providing the academic resources, facilities, and technical support that enabled the completion of this study.

My appreciation also goes to **Mr. Ioannis Sfyrakis**, my second reader, for his thoughtful suggestions and review of the project.

Finally, I am deeply thankful to my **family and friends** for their constant support, patience, and motivation throughout my academic journey. Their encouragement has been a source of strength and inspiration during every stage of this research.

Abstract

Customer churn is a major issue in the telecommunications industry. It affects revenue, customer retention, and long-term competitiveness. While machine learning models are commonly used to predict churn, many current methods focus on accuracy rather than transparency. This leads to black-box systems that business stakeholders find hard to trust and act on. This study suggests an explainable machine learning framework for predicting customer churn, using the Maven Telecom Customer Churn dataset.

The approach mixes descriptive analytics with supervised learning models, such as Logistic Regression, Random Forest, Gradient Boosting, LightGBM, and XGBoost. It predicts both the likelihood of customer churn and the churn category. To deal with class imbalance, we use SMOTE. We evaluate model performance using accuracy, F1-score, ROC-AUC, and cross-validation techniques. To improve interpretability, we apply Explainable Artificial Intelligence (XAI) methods, SHAP and LIME, to offer both global and local explanations of model predictions. The results show that ensemble-based models, especially XGBoost, provide strong predictive performance while also giving clear and actionable insights into the main factors driving churn. This supports ethical, understandable, and business-centered decision-making in managing telecom churn.

Keywords: Customer Churn Prediction, Telecommunications Industry, Machine Learning, Explainable Artificial Intelligence, SHAP, LIME, Ensemble Learning, XGBoost, Descriptive Analytics, Model Interpretability

Table Of Contents

List of Figures

List Of Tables

Chapter 1: Introduction

1.1 Background and Problem Context

Customer churn, which refers to the loss of customers either to competitors or due to discontinuing services, is a major financial challenge for the telecommunications industry. With fierce competition, low switching costs, and plenty of alternative service providers, it's much more cost-effective for telecom companies to retain existing customers than to acquire new ones (Verbeke et al., 2012). Because of this, telecom operators are turning more and more to data-driven insights to predict churn and develop proactive strategies to keep their customers.

Machine learning (ML) techniques have been widely adopted for churn prediction, leveraging historical customer data such as service usage, tenure, contract type, and billing information to estimate churn risk (Hastie et al., 2009). While these models often achieve high predictive accuracy, they are frequently criticised for operating as **black-box systems**, offering limited insight into the factors driving customer attrition. This lack of transparency restricts managerial trust, weakens business adoption, and limits the practical value of churn prediction systems (Molnar, 2022).

1.2 Customer Churn in the Telecommunications Domain

In the telecommunications context, customer churn refers to customers who terminate or fail to renew their subscription within a defined period. Churn may be **voluntary**, where customers actively switch providers due to dissatisfaction or competitive offers, or **involuntary**, resulting from billing issues or service termination.

Understanding churn solely as a binary outcome (churned vs non-churned) provides an incomplete picture of customer behaviour. From a business perspective, identifying the **underlying reasons** for churn—such as dissatisfaction with service quality, pricing concerns, or competitive pressure—is essential for designing effective and targeted retention interventions (Powers, 2011).

1.3 Motivation for Explainability and Descriptive Analytics

Traditional churn prediction research has primarily focused on optimising performance metrics such as accuracy, precision, and ROC-AUC. However, these metrics alone do not address the interpretability requirements of business stakeholders, who require transparency into how and why decisions are made.

Explainable Artificial Intelligence (XAI) has emerged as a solution to this challenge by providing interpretable explanations for complex machine learning models (Lundberg and Lee, 2017; Ribeiro et al., 2016). Techniques such as SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) enable both global and local interpretability, revealing how individual features contribute to predictions.

In addition to explainability, descriptive data analysis plays a crucial but often overlooked role in churn research. Many studies proceed directly from preprocessing to modelling, providing limited exploration of churn distributions, churn categories, or behavioural patterns (Kohavi, 1995). This project emphasises descriptive analysis as a foundational analytical layer, ensuring that predictive and explainable outputs are grounded in observable customer behaviour.

1.4 Problem Statement

Despite extensive research on telecom churn prediction, several limitations persist:

- Most studies focus exclusively on binary churn prediction, neglecting churn reasons.
- Explainability techniques are often applied superficially, without linking results to business interpretation.
- Descriptive analysis is underutilised, reducing contextual understanding of churn behaviour.
- Black-box models limit trust and real-world applicability.

As a result, telecom organisations are left with accurate predictions but insufficient insight into *why* customers churn or *how* to act on model outputs.

1.5 Research Aim

The primary aim of this study is:

To develop an explainable, multi-output machine learning framework that predicts both customer churn likelihood and churn reason, supported by descriptive analysis and interactive deployment.

The study leverages the Maven Telecom Customer Churn dataset, which uniquely includes churn categories and reasons, enabling richer modelling and interpretation than traditional binary datasets.

1.6 Research Questions

This research addresses the following questions:

1. How can machine learning models be extended to predict both customer churn and its underlying reason in the telecom sector?
2. How do SHAP and LIME enhance the interpretability and transparency of churn prediction models?
3. Which behavioural and business features most strongly influence different churn categories?
4. How can descriptive analytics and explainable predictions improve decision-making for customer retention strategies?

1.7 Research Hypothesis

This study hypothesises that integrating explainable AI techniques (SHAP and LIME) with a multi-output churn prediction framework and descriptive analysis will:

- Improve model transparency and interpretability,
- Reveal key business drivers behind customer churn,
- Maintain competitive predictive performance relative to traditional black-box models.

1.8 Objectives of the Study

The main goals of this research are to:

1. Review recent churn prediction literature (2021–2025) and identify gaps in explainability and descriptive insight.
2. Perform comprehensive descriptive analysis to explore churn distribution, categories, and reasons.
3. Preprocess and engineer features from the Maven Telecom Customer Churn dataset.
4. Develop and evaluate machine learning models for binary and multi-class churn prediction.
5. Apply SHAP and LIME to explain model behaviour at both global and individual levels.
6. Translate explainable outputs into business-relevant insights.
7. Deploy the final model using a Streamlit-based interactive web application for real-time prediction and analysis.
8. Assess the trade-offs between accuracy, interpretability, and ethical transparency.

1.9 Structure of the Report

This report is organised as follows:

- **Chapter 1 – Introduction**: Introduces the research context, motivation, aims, and objectives.
- **Chapter 2 – Literature Review**: Reviews existing churn prediction research and identifies key gaps.
- **Chapter 3 – Methodology**: Describes the dataset, preprocessing, modelling framework, XAI techniques, and tools used.
- **Chapter 4 – Exploratory Data Analysis**: Presents descriptive statistics and visual analysis of churn behaviour.
- **Chapter 5 – Results and Discussion**: Evaluates model performance, explainability outcomes, and deployment results.
- **Chapter 6 – Conclusion and Future Work**: Summarises findings, contributions, limitations, and future research directions.

Chapter 2 – Literature Review

2.1 Introduction

Customer churn prediction has been a central research topic in data-driven business analytics, particularly within the telecommunications sector where customer retention has a direct and measurable impact on revenue and long-term profitability (Verbeke et al., 2012). Over the period 2021–2025, advances in machine learning (ML) and ensemble learning have significantly improved churn prediction accuracy, with many studies reporting performance exceeding 80% accuracy on benchmark datasets (Chen and Guestrin, 2016; Zhang et al., 2020).

Despite these advances, recent literature reveals critical methodological gaps. Most churn prediction studies prioritise predictive accuracy while giving limited attention to explainability, descriptive insight, and business interpretability (Molnar, 2022). As a result, many high-performing models function as black boxes, providing limited support for managerial decision-making. This chapter critically reviews recent churn prediction literature and identifies how the current research addresses these gaps by integrating descriptive analytics, multi-output learning, and explainable artificial intelligence (XAI) within a unified framework.

2.2 Overview of Customer Churn Prediction Studies (2021–2025)

Telecom churn prediction research typically relies on structured customer data, including demographic attributes, service usage, billing information, and contract characteristics. Commonly applied algorithms include Logistic Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), Gradient Boosting, and XGBoost (Breiman, 2001; Chen and Guestrin, 2016).

Recent studies frequently employ ensemble learning techniques to optimise evaluation metrics such as accuracy, F1-score, and ROC-AUC (Zhang et al., 2020). While these approaches demonstrate strong predictive performance, they often lack interpretability and contextual explanation. A substantial proportion of this literature relies on the IBM Telco Customer Churn dataset, which provides only a single binary churn label (Yes/No). Consequently, these models identify *who* is likely to churn but fail to explain *why* churn occurs (Verbeke et al., 2012).

This limitation restricts practical business applicability, as churn prediction without explanation does not directly inform retention strategies or operational interventions.

2.3 Prediction Goal: From Binary to Multi-Output Learning

A dominant limitation in existing churn prediction research is the binary-only prediction objective. Most studies focus exclusively on estimating the probability of churn, addressing the question: *Will the customer churn?* (Hastie et al., 2009).

However, churn is a multi-dimensional phenomenon, influenced by factors such as service dissatisfaction, pricing concerns, and competitive pressure. Very few studies attempt to predict both churn occurrence and churn cause, largely due to dataset constraints (Molnar, 2022).

This research adopts a multi-output learning approach, simultaneously predicting:

1. Customer churn likelihood (Churned vs Stayed), and
2. Churn category or reason (e.g., Dissatisfaction, Competitor, Other).

Such dual-level prediction aligns analytical outputs with real-world business needs, enabling targeted and reason-specific retention strategies. The use of the Maven Telecom Customer Churn dataset, which includes explicit churn categories and reasons, enables this methodological advancement and addresses a notable gap in the literature.

2.4 Role of Descriptive Analysis in Churn Understanding

While predictive modelling dominates churn research, descriptive analysis remains underutilised in recent studies. Many papers move directly from data preprocessing to model training, providing minimal exploration of churn distributions, churn categories, or churn reasons (Kohavi, 1995).

Descriptive analytics plays a crucial role in:

- Understanding churn prevalence and imbalance,
- Identifying dominant churn categories,
- Revealing business-driven churn patterns prior to modelling.

This study places descriptive analysis as a foundational step, not merely a preliminary task. By analysing customer status distribution, churn categories, churn reasons, and cross-tabulated relationships, the research establishes contextual understanding that informs both feature engineering and interpretability. This emphasis on descriptive insight represents a practical and methodological contribution that is largely absent in prior churn prediction research.

2.5 Explainable AI in Churn Prediction Models

Explainable Artificial Intelligence (XAI) has gained increasing attention as regulatory, and business demands for transparency grow (Molnar, 2022). Techniques such as SHAP (Lundberg and Lee, 2017) and LIME (Ribeiro et al., 2016) have been applied in recent churn studies to identify feature importance.

However, existing research often applies XAI in a superficial and technical manner, presenting feature rankings without linking explanations to business meaning. For example, identifying “Monthly Charges” or “Tenure” as important predictors does not inherently explain customer dissatisfaction or competitive switching.

This study advances beyond this limitation by explicitly linking XAI outputs to churn categories and business causes, transforming model explanations into actionable insights. SHAP provides global and local feature attribution, while LIME validates individual predictions, ensuring consistency and trustworthiness across explanation levels.

2.6 Dataset Limitations in Prior Studies

A recurring constraint in churn prediction literature is dataset simplicity. The IBM Telco dataset, despite its popularity, lacks churn reasons, limiting research to binary outcomes (Verbeke et al., 2012).

This dataset limitation prevents:

- Multi-output modelling,
- Cause-specific churn analysis,
- Fine-grained customer segmentation.

The Maven Telecom Customer Churn dataset overcomes these limitations by including Churn Category and Churn Reason, enabling richer analysis and more realistic modelling. This dataset choice represents a deliberate methodological improvement over prior study.

2.7 Business Interpretability and Decision Support

Many existing studies evaluate models solely through technical metrics, neglecting business relevance (Powers, 2011). High accuracy alone does not guarantee managerial usefulness.

This research bridges that gap by:

- Translating descriptive insights into churn drivers,
- Mapping SHAP and LIME explanations to business actions,
- Supporting targeted retention strategies.

By embedding interpretability throughout the modelling pipeline, the study aligns predictive analytics with operational decision-making.

2.8 Ethical and Professional Considerations in Prior Research

Recent literature highlights the ethical risks of black-box decision systems, particularly in customer-facing applications (Molnar, 2022). Many churn studies fail to address transparency, accountability, or fairness.

This project adopts an ethical-by-design approach, using XAI to ensure that predictions are explainable, auditable, and justifiable, supporting responsible AI deployment in business environments.

2.9 Summary of Identified Research Gaps

The literature review identifies six consistent gaps:

1. Over-reliance on binary churn prediction
2. Limited use of descriptive analytics
3. Superficial application of XAI
4. Dataset constraints preventing causal analysis
5. Absence of multi-output modelling
6. Weak ethical and business grounding

2.10 How This Study Addresses the Gap

This research fills these gaps by:

- Employing a rich dataset with churn reasons,
- Implementing multi-output learning,
- Integrating descriptive analysis as a core component,
- Applying SHAP and LIME for business-level explanations,
- Supporting ethical and interpretable decision-making.

2.11 Conclusion

The reviewed literature demonstrates that while churn prediction accuracy has improved, explainability and descriptive understanding remain underdeveloped. By combining descriptive analytics, multi-output modelling, and explainable AI, this study provides a holistic and business-aligned churn prediction framework, contributing both academically and practically to the telecommunications domain.

Chapter 3 – Methodology

3.1 Introduction

This chapter outlines the research design, dataset, data preprocessing procedures, modelling framework, and explainable AI methods used in this study. The research follows a quantitative, experimental approach that combines predictive modelling with interpretability techniques to identify both the likelihood and the reason for customer churn. The methodology was designed to ensure reliability, transparency, and business applicability, aligning with the research aim of building an explainable multi-output machine learning framework using the Maven Telecom Customer Churn dataset(😦 📡 Telecom Customer Churn Prediction).

3.2 Research Design and Approach

The project adopts a supervised machine learning approach consisting of two interconnected stages:

1. **Binary Classification Model** – predicts whether a customer will churn or not.
2. **Multi-Class Classification Model** – identifies the category or reason for churn (e.g., Dissatisfaction, Competitor, Other).

The study integrates Explainable AI (XAI) methods — SHAP and LIME — to interpret model predictions at both the global and individual levels. This dual approach enhances both predictive accuracy and interpretability, producing actionable insights for telecom businesses.

A quantitative experimental design was chosen because it allows for systematic evaluation of different models using measurable performance metrics such as Accuracy, Precision, Recall, F1-score, and ROC-AUC. The experimental framework includes model comparison, validation, and interpretability assessment using visualization and statistical evaluation.

Figure 1: ***Customer Churn Prediction Research Workflow chart***

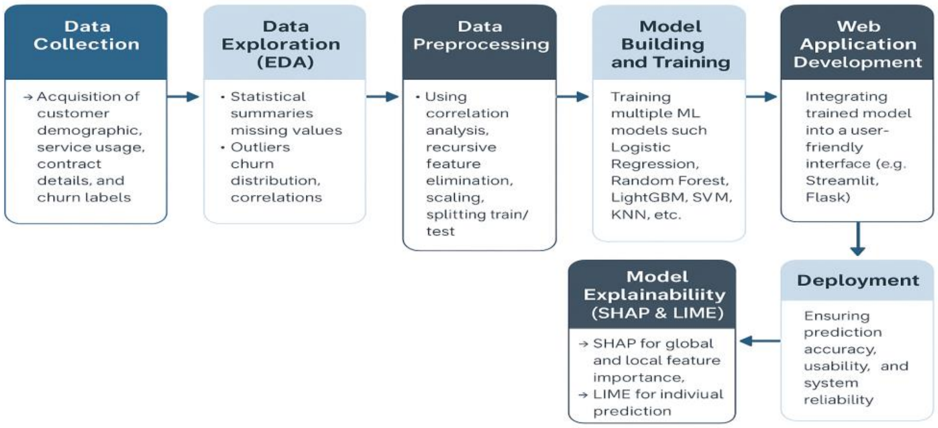

Figure 1 illustrates the complete workflow of the customer churn prediction process used in this study. It starts with gathering data and performing exploratory analysis to better understand customer behaviour and churn patterns. The data is then pre-processed, which includes cleaning, encoding, feature selection, and splitting into training and test sets. Several machine learning models are trained and evaluated to predict customer churn. Model explanation methods like SHAP and LIME are used to interpret the predictions. Lastly, the trained model is deployed through a web application for practical business use.

3.3 Dataset Description

The research utilises the Maven Telecom Customer Churn dataset, a publicly available dataset hosted on Kaggle by Maven Analytics. It contains detailed records of telecom customers, including demographic details, account information, subscription plans, service usage, billing patterns, and churn behaviour.

Key features include:

- **Customer Status** – indicates whether the customer is *Churned*, *Stayed*, or *Joined* (target for binary prediction).
- **Churn Category** – represents the main reason for churn such as *Dissatisfaction*, *Competitor*, *Other Reasons* (target for multi-class prediction).
- **Churn Reason** – provides qualitative explanations (e.g., “Attitude of Support Personnel”).
- Tenure Months, Monthly Charges, Total Charges, Contract Type, Internet Type, Payment Method, and Customer Satisfaction Score – predictor variables.

The dataset is fully anonymised and ethically appropriate for academic use. It consists of approximately 7,000–8,000 customer records and 50+ attributes, allowing comprehensive exploration of behavioural and business patterns related to churn.

3.4 Data Collection and Ethical Considerations

The dataset was sourced from a verified public repository (Kaggle) and does not contain any personally identifiable information (PII). It complies with Teesside University's ethical data-handling standards. No human participants were involved in the study. All experiments were conducted locally using Python in Jupyter Notebook. File handling and processing followed standard data protection guidelines, ensuring data anonymity and reproducibility.

3.5 Data Preprocessing

To ensure data quality and model readiness, several preprocessing steps were performed as outlined below:

3.5.1 Handling Missing and Duplicate Values

- Missing values were identified using df.isnull().sum().
- Depending on the type of variable, missing numerical values were replaced with the mean or median, while categorical variables were filled in with the most frequent value (mode).
- Duplicate entries were removed using the drop_duplicates() method to maintain dataset integrity.

3.5.2 Data Type Correction

Certain numeric fields such as *Total Charges* or *Tenure Months* were stored as strings; these were converted to numeric formats using pd.to_numeric() to ensure compatibility with ML algorithms.

3.5.3 Feature Selection

Irrelevant columns (e.g., *Customer ID*, *Zip Code*, *Latitude*, *Longitude*) were dropped as they do not contribute to prediction. Feature importance analysis during model evaluation further refined the most impactful variables.

3.5.4 Encoding Categorical Variables

Categorical features (e.g., *Contract Type*, *Internet Service*, *Payment Method*) were encoded using Label Encoding or One-Hot Encoding, depending on the algorithm's requirement. This step transformed text labels into numeric representations suitable for model ingestion.

3.5.5 Feature Scaling

Continuous variables such as Monthly Charges, Tenure Months, and Total Charges were normalized using StandardScaler from Scikit-learn to ensure consistent feature magnitude and reduce algorithmic bias.

3.5.6 Handling Class Imbalance

Since the number of churned customers was smaller compared to non-churned, class imbalance was mitigated using SMOTE (Synthetic Minority Oversampling Technique) to generate synthetic examples for the minority class, improving model fairness and stability.

3.6 Exploratory Data Analysis (EDA)

Before modelling, a detailed exploratory analysis was conducted to understand variable distributions and relationships. Key visualizations included:

- Churn rate distribution showing overall churn percentage.
- Correlation heatmap between numerical features.
- Churn by contract type and payment method, revealing behavioural trends.
- Revenue impact of churned customers, identifying high-value loss segments. EDA insights informed both feature selection and hypothesis formulation, highlighting that short tenure, month-to-month contracts, and high monthly charges correlated strongly with churn.

3.7 Model Development Framework (With Necessary Mathematical Formulation)

To ensure robustness and comprehensive evaluation, three widely used supervised machine learning algorithms were implemented and compared in this study. These models represent increasing levels of complexity, ranging from linear classification to ensemble-based methods capable of capturing non-linear relationships in customer churn data.

Two modelling pipelines were developed:

- **Model 1: Binary Classification** – predicts customer status (*Churned vs Stayed*)
- **Model 2: Multi-Class Classification** – predicts churn category (*Competitor, Dissatisfaction, Other*)

Each model was trained using 80% of the dataset and evaluated on the remaining 20% to assess generalisation performance.

3.7.1 Logistic Regression

Logistic Regression is a type of model used to predict the likelihood that a customer will churn, based on a set of input factors. It works by taking a weighted sum of these factors and then applying the logistic (or sigmoid) function, which converts the result into a probability value between 0 and 1. This helps the model estimate how likely it is for a customer to leave.

The predicted probability of churn is defined as:

$$P(y = 1 \mid X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)}}$$

where:

- $X = (x_1, x_2, \ldots, x_n)$represents the feature vector
- $\beta_0$is the intercept
- $\beta_i$are the model coefficients

A customer is classified as churned if the predicted probability exceeds a defined threshold. Logistic Regression assumes a linear relationship between input features and the log-odds of the target variable, which limits its ability to capture complex non-linear patterns. Nevertheless, its simplicity and interpretability make it a suitable baseline model.

3.7.2 Random Forest Classifier

Random Forest is an ensemble learning algorithm that constructs multiple decision trees during training and aggregates their predictions to produce a final output. Each decision tree is trained on a random subset of the training data and a random subset of features, which reduces variance and improves generalisation.

The final prediction of a Random Forest classifier is made by taking a vote, with the most common prediction winning:

$$\hat{y} = \text{mode}\{h_1(X), h_2(X), \ldots, h_T(X)\}$$

Where:

- $ht(X)$ represents the prediction made by the t decision tree.
- T is the total number of trees in the forest.

Random Forest can model complex non-linear relationships and interactions between features without requiring explicit assumptions about data distribution. Additionally, it provides internal feature importance measures based on impurity reduction, aiding interpretability.

3.7.3 XGBoost Classifier

XGBoost (Extreme Gradient Boosting) is an advanced algorithm that builds decision trees one after another. Each new tree is trained to fix the mistakes made by the previous ones by minimizing a regularized objective function.

The general objective function optimised by XGBoost is given by:

$$\mathcal{L} = \sum_{i=1}^{N} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

where:

- $l(y_i, \hat{y}_i)$is the loss function measuring prediction error
- $\Omega(f_k)$is a regularisation term controlling model complexity
- $f_k$represents the k^{th}decision tree

By incorporating regularisation and second-order optimisation techniques, XGBoost effectively balances bias and variance, resulting in strong predictive performance and improved generalisation. Its ability to model subtle patterns and non-linear decision boundaries makes it particularly suitable for telecom churn prediction tasks.

Training and Evaluation Setup

All models were trained using stratified train–test splits to preserve class distribution. For multi-class prediction, the models were configured using appropriate multi-class strategies. Performance evaluation was conducted using standard classification metrics and confusion matrices to assess prediction quality and misclassification behaviour.

3.8 Explainable AI Integration (XAI Layer)

To ensure transparency and interpretability, explainable AI tools were applied to the final models.

3.8.1 SHAP (SHapley Additive Explanations)

SHAP was used to determine how each feature contributes to individual predictions.

- **Global SHAP Analysis:** Identified overall feature importance (e.g., *Tenure*, *Monthly Charges*, *Contract Type*).
- **Local SHAP Analysis:** Provided customer-level explanations showing why a specific individual is predicted to churn.

3.8.2 LIME (Local Interpretable Model-Agnostic Explanations)

LIME was used to validate SHAP insights by locally approximating complex model decisions with simpler interpretable models. It provided case-by-case explanations, supporting consistency between local and global interpretations.

Together, SHAP and LIME established a transparent explanation framework, bridging the gap between model output and business understanding.

3.9 Model Evaluation Metrics

Model performance was evaluated using quantitative metrics designed to assess predictive accuracy while accounting for class imbalance. These metrics provide a balanced evaluation of model effectiveness, particularly for churn prediction where misclassification costs are asymmetric.

Accuracy measures the overall correctness of predictions and is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

where *TP* denotes true positives, *TN* true negatives, *FP* false positives, and *FN* false negatives.

Precision evaluates the proportion of correctly identified churned customers among all customers predicted as churned:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall measures how well the model can accurately identify customers who have churned:

$$\text{Recall} = \frac{TP}{TP + FN}$$

The F1-score strikes a balance between precision and recall, making it especially useful for datasets where the classes are imsbalanced:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Along with these metrics, the Receiver Operating Characteristic – Area Under the Curve (ROC-AUC) was used to assess how well the model distinguishes between different classes at various classification thresholds. A higher AUC value indicates better separation between churn and non-churn customers.

Confusion matrices were also analysed to visualise classification outcomes and identify patterns of misclassification, particularly false negatives, which are critical in churn prediction tasks.

3.10 Tools and Technologies

The project was developed using the following technologies:

- **Programming Language:** Python 3.10
- **Libraries:** Pandas, NumPy, Scikit-learn, XGBoost, SHAP, LIME, Matplotlib, Seaborn
- **IDE:** Jupyter Notebook
- **Deployment Framework: Streamlit** (for web application)
- **Version Control:** GitHub
- **Visualization:** Matplotlib & Seaborn for EDA; SHAP & LIME for interpretability.

3.11 Deployment and Application Framework

To support practical applicability and real-world usage, the proposed churn prediction framework was designed with deployment considerations in mind. The final trained machine learning model was incorporated into an interactive web application built using Streamlit.

Streamlit was chosen because of its lightweight design and its ability to quickly deploy data-driven applications. The framework allows users to input customer attributes and obtain churn predictions along with probability estimates generated by the trained model.

From a methodological perspective, this deployment layer ensures that the predictive system is not limited to offline analysis but can be operationalised as a decision-support tool for business stakeholders. The integration of a web interface aligns with the project's objective of developing a usable and explainable machine learning solution rather than a purely theoretical model.

3.12 Ethical, Legal, and Professional Considerations

The project strictly adheres to ethical data science principles. The dataset is anonymised, ensuring no privacy or consent violations. Model explainability serves an ethical purpose by preventing algorithmic bias and enhancing user trust. All tools and libraries used are open-source, and no commercial or proprietary data was accessed. Transparency, reproducibility, and fairness were maintained throughout the study.

Chapter 4: Implementation

4.1 Overview of Implementation

This chapter presents the practical implementation of the customer churn prediction framework described in Chapter 3. It focuses on the application of data analysis techniques, feature transformation processes, and machine learning models to extract meaningful insights and build predictive solutions from the telecom dataset.

The chapter begins with exploratory data analysis (EDA) to examine customer behaviour patterns, class distributions, and relationships among key variables. This is followed by data transformation and feature engineering, where raw features are encoded, scaled, and enhanced to improve model learning capability. Finally, the chapter details the implementation and optimisation of multiple machine learning models, including training strategies, validation approaches, and hyperparameter tuning to identify the most effective predictive model.

Overall, this chapter bridges the methodological design outlined in Chapter 3 with the empirical results presented in Chapter 5, demonstrating how the proposed techniques are applied in practice to address the research objectives.

4.2 Statistical Summary and Exploratory Data Analysis

The purpose of this section is to explore the structure of the dataset and identify key patterns related to customer churn behaviour. Exploratory Data Analysis (EDA) provides an initial understanding of customer characteristics, service usage, and contractual factors that may influence churn. Insights obtained from this analysis inform feature engineering decisions and guide model development in later stages.

4.2.1 Dataset Overview

The dataset comprises 7,043 customer records with over 50 demographic, service, and billing-related features. The primary target variable, Customer Status, consists of three classes: Stayed, Churned, and Joined.

The class distribution is imbalanced, with 4,720 customers staying, 1,869 customers churning, and 454 customers joining the service. This imbalance indicates that churned customers form a smaller but critical segment of the dataset, which can influence model learning and evaluation. As a result, appropriate strategies such as class balancing and the use of robust evaluation metrics are necessary to ensure fair and reliable model performance.

4.2.2 Customer Status Distribution

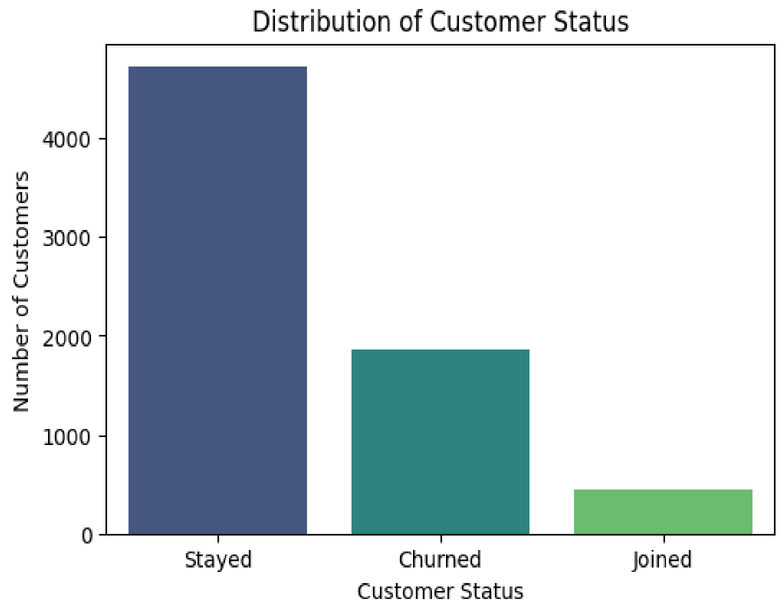

*Figure 2**: Distribution of Customer Status***

Figure 2 presents the overall distribution of customer status within the dataset. The majority of customers belong to the Stayed category, while a substantial proportion have Churned, and a smaller fraction represent Joined customers. Although churn events occur less frequently than retention, their presence remains significant due to their direct impact on business revenue. This distribution highlights the importance of designing predictive models that effectively identify churn despite class imbalance.

4.2.3 Churn by Contract Type

Figure 3 illustrates customer churn behaviour across different contract types. Customers on Month-to-Month contracts exhibit the highest churn rates, indicating that flexible contractual arrangements are associated with increased churn risk. In contrast, One-Year and Two-Year contracts show higher retention, with Two-Year contracts demonstrating the lowest churn levels. These findings suggest that longer-term contractual commitments play a crucial role in improving customer retention.

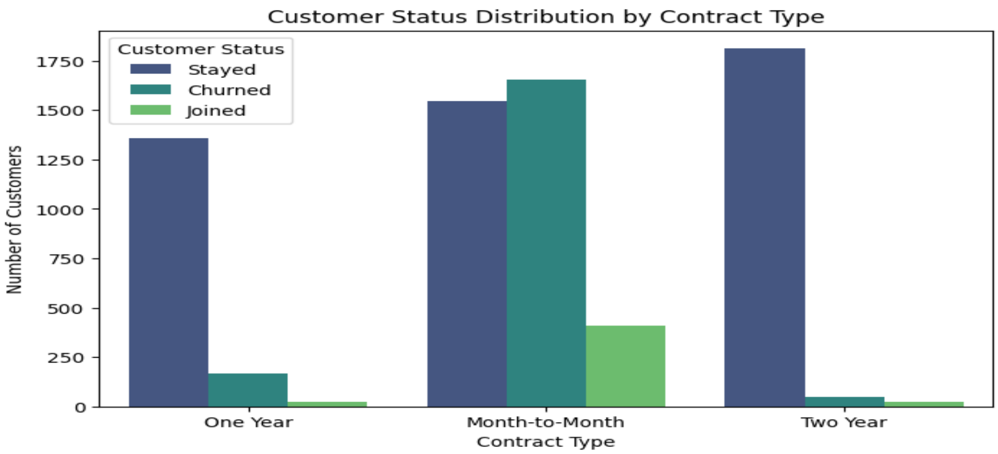

Figure 3: ***Customer status Distribution by Contract Type***

4.2.4 Churn Across Categorical Features

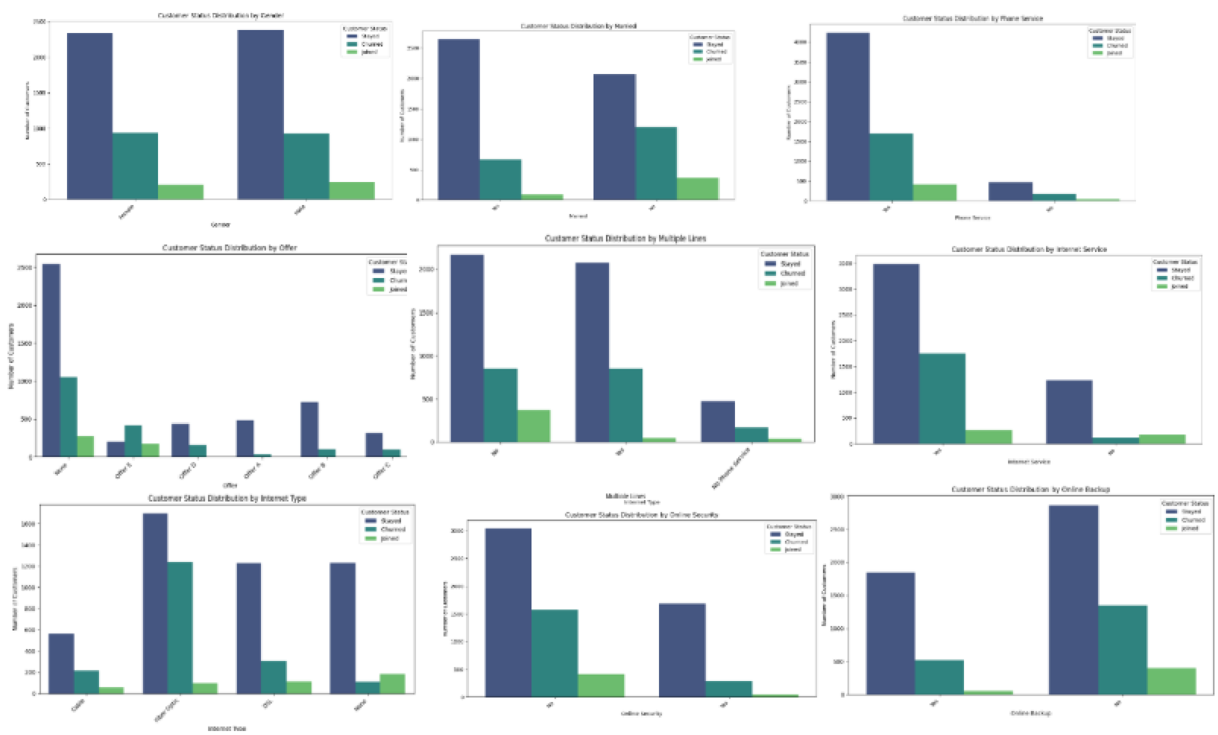

Figure 4: ***Customer Status Distribution Across All Categorical Features***

Figure 4 shows churn patterns across various categorical features, including gender, marital status, promotional offers, service subscriptions, and internet type. Gender does not appear to influence churn behaviour significantly; however, unmarried customers show higher churn rates compared to married customers. Customers without promotional offers and those on flexible contracts exhibit increased churn, highlighting the impact of incentives and contract stability.

Additionally, customers lacking services such as online security, online backup, and phone service demonstrate elevated churn levels. Analysis of internet type indicates that Fiber Optic customers experience higher churn, suggesting potential concerns related to pricing or service quality. Overall, these observations indicate that service engagement, contract type, and value-added features strongly influence customer retention.

4.2.5 Correlation Analysis of Numerical Features

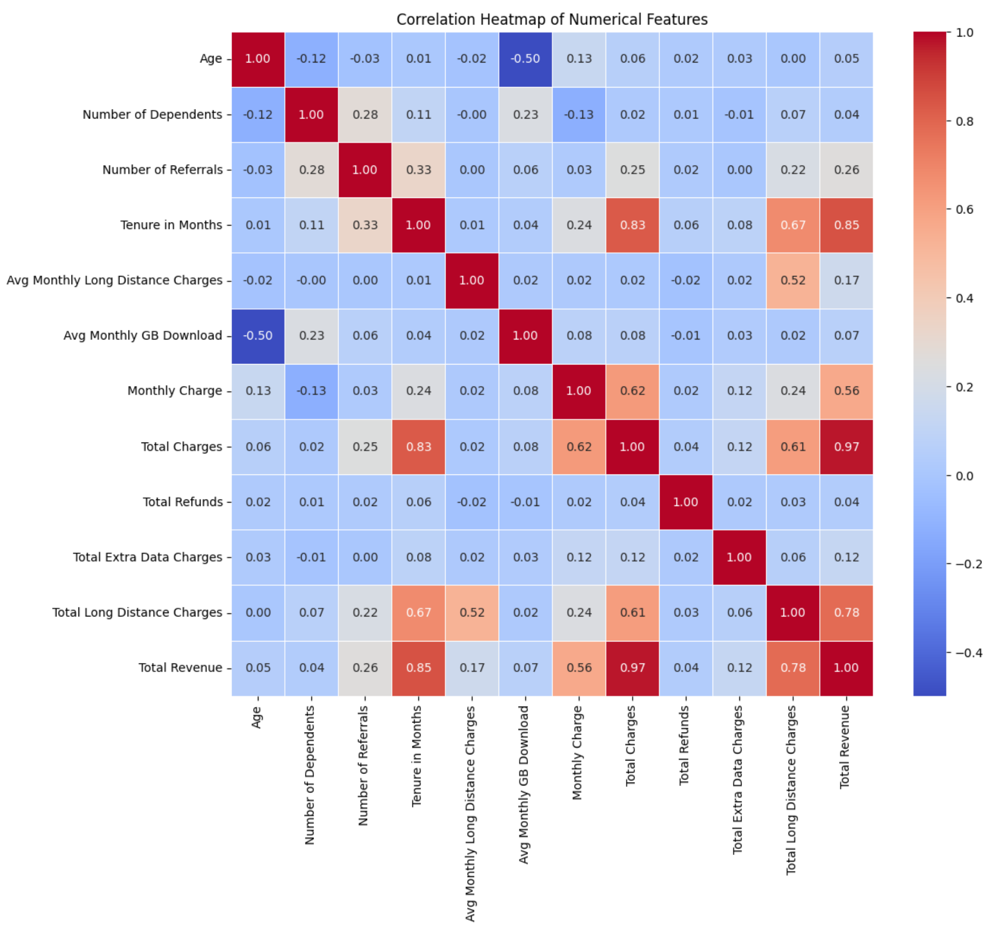

Figure 5: ***Correlation Heatmap of Numerical Features***

Figure 5 presents a correlation heatmap showing relationships among numerical features. Strong positive correlations are observed between Monthly Charges, Total Charges, and

Total Revenue, indicating that these financial attributes increase together. Tenure shows a moderate positive correlation with revenue-related features, reflecting higher lifetime value for long-term customers.

Most other numerical variables exhibit weak correlations, suggesting limited linear dependency and reinforcing the need for machine learning models capable of capturing non-linear relationships. These insights support the selection of advanced ensemble models in subsequent modelling stages.

4.2.6 Descriptive Analysis of Churn Categories and Reasons

A descriptive analysis of churn categories and churn reasons was conducted to identify the primary drivers of customer churn and to provide contextual understanding for the predictive models developed later in this study.

An examination of Customer Status shows that most customers remained with the service (4,720), while 1,869 customers churned and 454 customers joined. Although churn represents a minority class, its proportion is substantial and highlights the importance of understanding the underlying causes of customer attrition.

Analysis of the Churn Category variable indicates that churn is largely driven by competitor-related factors, which account for the largest share of churned customers. Categories associated with dissatisfaction, attitude, and price also contribute significantly, suggesting that both external competition and internal service-related issues influence churn behaviour. Customers who remained or joined the service are predominantly associated with the *No Churn* category, confirming the relevance of churn categories in distinguishing churned customers.

A detailed assessment of the top churn reasons among churned customers provides further insight into specific drivers of attrition. The most frequently reported reasons include competitors offering better devices, competitors providing more attractive offers, and negative customer support experiences. Other notable factors include uncertainty regarding the reason for churn, higher data allowances offered by competitors, and perceptions of high pricing. These findings emphasise the combined impact of competitive market dynamics and customer experience on churn decisions.

Cross-tabulation between Customer Status and both Churn Category and Churn Reason confirms that these variables are strongly associated with churned customers, reinforcing their analytical and business relevance. Overall, the descriptive findings demonstrate that customer churn is influenced by identifiable and interpretable factors rather than occurring randomly.

The insights derived from this analysis informed feature selection, feature engineering, and the design of the multi-class churn category prediction model, providing a strong foundation for the predictive and explainable machine learning framework presented in the following chapters.

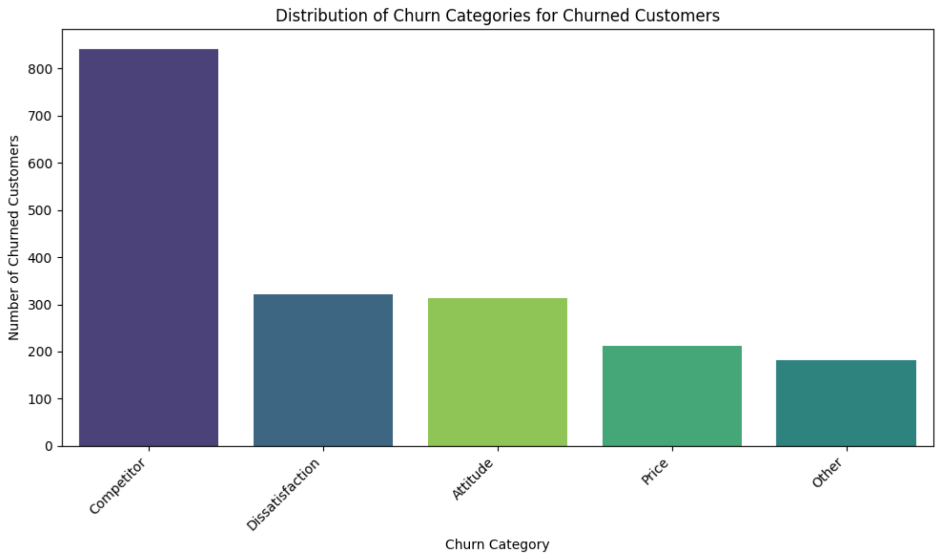

Figure 6: ***Distribution of Churn Categories Among Churned Customers***

Figure 6 presents the distribution of churn categories for customers who have churned from the telecom service. The largest proportion of churn is attributed to competitor-related reasons, indicating strong competitive pressure within the market. Categories such as dissatisfaction and attitude-related issues also contribute notably to customer churn, highlighting the influence of service quality and customer experience. In contrast, price and other reasons account for a smaller share of churn cases. Overall, this visualization emphasises that both external competition and internal service factors play a significant role in driving customer churn.

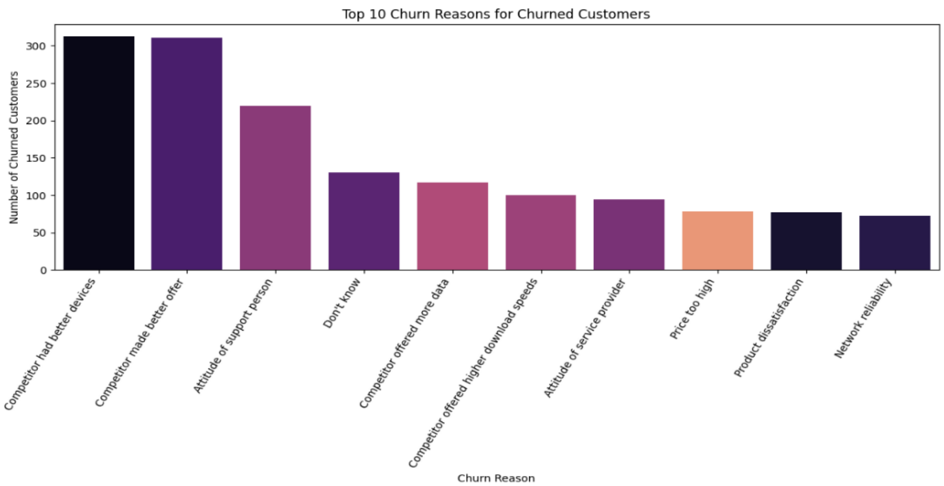

Figure 7: ***Top 10 Churn Reasons Among Churned Customers***

Figure 7 displays the ten most frequently reported reasons for customer churn. The most dominant factors are competitor-related offerings, such as better devices and more attractive plans, indicating strong market competition. Service-related issues, including the attitude of support personnel and service provider behaviour, also contribute substantially to churn. Pricing concerns and product dissatisfaction appear less frequently but remain relevant. Overall, the figure highlights that both competitive pressure and customer service quality are key drivers of churn, providing important insights for targeted retention strategies.

4.3 Data Transformation and Feature Engineering

The purpose of this section is to transform the raw dataset into a structured and model-ready format. Based on insights obtained from the exploratory data analysis, appropriate transformation and feature engineering techniques were applied to improve model learning, ensure numerical stability, and capture meaningful customer behaviour patterns.

4.3.1 Feature Encoding

The dataset contains multiple categorical variables, such as contract type, internet service, payment method, marital status, and service subscriptions, which cannot be directly processed by most machine learning algorithms. To address this, categorical features were converted into numerical representations using label encoding and one-hot encoding, depending on the nature of the variable.

Binary categorical variables were encoded using label encoding, while nominal variables with multiple categories were transformed using one-hot encoding. This approach preserves categorical information while preventing unintended ordinal relationships, ensuring compatibility with both linear and tree-based models.

```
Categorical columns to encode: ['Gender', 'Married', 'Offer', 'Phone Service', 'Multiple Lines', 'Internet Service', 'Internet Type', 'Online Security', 'Online Backup', 'Device Protection Plan',
```

	Age	Number of Dependents	Number of Referrals	Tenure in Months	Avg Monthly Long Distance Charges	Avg Monthly GB Download	Monthly Charge	Total Charges	Total Refunds	Total Extra Data Charges	...	Premium Tech Support_Yes	Streaming TV_Yes	Streaming Movies_Yes	Streaming Music_Yes	Unlimited Data_Yes	Contract_One Year	Contract_Two Year	Paperless Billing_Yes
0	37	0	2	9	42.39	16.0	65.6	593.30	0.00	0	...	True	True	False	False	True	True	False	True
1	46	0	0	9	10.69	10.0	-4.0	542.40	38.33	10	...	False	False	True	True	False	False	False	False
2	50	0	0	4	33.65	30.0	73.9	280.85	0.00	0	...	False	False	False	False	True	False	False	True
3	78	0	1	13	27.82	4.0	98.0	1237.85	0.00	0	...	False	True	True	False	True	False	False	True
4	75	0	3	3	7.38	11.0	83.9	267.40	0.00	0	...	True	True	False	False	True	False	False	True

5 rows × 45 columns

```
Shape of the encoded DataFrame: (7043, 45)
```

Figure 8: ***Sample of Dataset After Categorical Feature Encoding***

Figure 8 shows a sample of the dataset after categorical feature encoding. Categorical variables such as contract type, internet services, streaming services, and billing options have been transformed into numerical representations using encoding techniques. The resulting dataset consists of 7,043 records and 45 features, making it suitable for input into machine learning models. This transformation enables algorithms to effectively process categorical information while preserving relevant customer characteristics for churn prediction.

4.3.2 Feature Scaling

Numerical features such as **Tenure in Months, Monthly Charges, Total Charges, and Total Revenue** exhibit varying magnitudes, which can bias distance-based and gradient-based algorithms. To address this issue, **feature scaling** was applied to standardise numerical variables.

Standardisation was performed using **StandardScaler**, which transforms features to have zero mean and unit variance. This process improves convergence speed, ensures consistent feature contribution, and enhances the performance of algorithms sensitive to feature scale, such as Logistic Regression and Support Vector Machines.

4.3.3 Feature Engineering

To capture complex relationships and enhance predictive power, additional features were engineered using domain knowledge and EDA insights. **Interaction features**, such as the relationship between customer tenure and monthly charges, were created to model non-linear customer behaviour.

Furthermore, **aggregate service-based features** were constructed by combining multiple service subscription indicators into a single variable representing overall service engagement. This feature provides a compact representation of customer involvement across services.

Finally, **business-driven transformations** were applied to highlight customer value and risk patterns, such as combining usage and billing attributes to better represent customer lifetime contribution. These engineered features enabled the models to distinguish more effectively between churned and retained customers.

	Tenure in Months	Monthly Charge	Tenure_Monthly_Charge	Number of Referrals	Referrals_Tenure	Total Services	Phone Service	Multiple Lines
0	9	65.6	590.4	2	18	6	Yes	No
1	9	-4.0	-36.0	0	0	5	Yes	Yes
2	4	73.9	295.6	0	0	4	Yes	No
3	13	98.0	1274.0	1	13	7	Yes	No
4	3	83.9	251.7	3	9	5	Yes	No

Figure 9: ***Sample of Engineered Features in the Transformed Dataset***

Figure 9 shows a sample of the dataset after feature engineering, including interaction and aggregate features designed to better capture customer behaviour for churn prediction.

4.4 Model Implementation

This section describes the implementation of machine learning models used to predict customer churn. The objective is to build reliable predictive models that can learn from customer data and generalise well to unseen instances. Both baseline and ensemble models were implemented to enable performance comparison and identify the most effective approach.

4.4.1 Train–Test Split Strategy

The dataset was divided into training and testing subsets to evaluate model performance on unseen data. An 80:20 train–test split was adopted, ensuring that the majority of data was used for model training while reserving a sufficient portion for independent testing.

To preserve the original class distribution of the target variable, stratified sampling was applied during the splitting process. This approach ensures that each subset contains proportional representations of Stayed, Churned, and Joined customers, which is particularly important given the observed class imbalance. Stratification helps prevent biased performance estimates and improves the reliability of evaluation results.

```
Shape of X_train_resampled: (11328, 20)
Shape of y_train_resampled: (11328,)

Value counts for resampled y_train_encoded:
[3776 3776 3776]

Value counts for resampled y_train_encoded with labels:
{'Churned': np.int64(3776), 'Joined': np.int64(3776), 'Stayed': np.int64(3776)}
```

Figure 10: ***Class Distribution After SMOTE Resampling***

Figure 10 shows the class distribution of the training dataset after applying the Synthetic Minority Over-sampling Technique (SMOTE). The resampled training set contains 11,328 instances, with an equal number of samples (3,776 each) for the Churned, Joined, and Stayed classes. This balanced distribution ensures that the machine learning models are not biased toward the majority class and can learn decision boundaries more effectively.

4.4.2 Baseline Model Implementation

Logistic Regression was implemented as a baseline model to provide a simple and interpretable benchmark for comparison. As a linear classifier, Logistic Regression offers transparency in decision-making and serves as a reference point to assess the added value of more complex models.

The model was trained using the transformed feature set and evaluated using standard classification metrics. Although its predictive capacity is limited in capturing non-linear relationships, it provides valuable insight into fundamental churn drivers and establishes a baseline performance level.

4.4.3 Ensemble Model Implementation

To capture complex, non-linear relationships within the data, several ensemble learning models were implemented.

The Random Forest Classifier was employed as a bagging-based ensemble method that combines multiple decision trees to reduce variance and improve robustness. Its ability to handle mixed data types and model feature interactions makes it well-suited for churn prediction tasks.

Additionally, gradient boosting models, including XGBoost and LightGBM, were implemented due to their proven effectiveness in structured tabular data. These models iteratively improve predictions by focusing on previously misclassified instances, resulting in strong generalisation performance and high predictive accuracy. Their flexibility and scalability make them particularly suitable for large, feature-rich datasets such as telecom customer data.

4.5 Model Optimisation

The purpose of model optimisation is to enhance predictive performance and improve generalisation to unseen data. After initial model implementation, optimisation techniques were applied to refine model parameters and reduce the risk of overfitting. This process ensures that the selected models achieve reliable and stable performance across different data splits.

4.5.1 Hyperparameter Tuning

To identify the most effective model configurations, hyperparameter tuning was performed using systematic search strategies. For computational efficiency and broader parameter exploration, RandomizedSearchCV was employed to sample combinations of hyperparameters across predefined ranges. This approach enables efficient optimisation while avoiding the exhaustive computational cost associated with GridSearchCV.

Key hyperparameters, such as the number of trees, tree depth, learning rate, and regularisation parameters, were tuned for ensemble models. The tuning process was guided by cross-validation performance, ensuring that selected parameter values improved generalisation rather than fitting noise in the training data.

4.5.2 Cross-Validation Strategy

To obtain robust and unbiased performance estimates, Stratified K-Fold cross-validation was applied during model optimisation. This method divides the dataset into multiple folds while maintaining the original class distribution in each fold, which is essential given the observed class imbalance.

Cross-validation ensures that each observation is used for both training and validation, providing a more reliable assessment of model performance. By combining hyperparameter tuning with stratified cross-validation, the optimisation process enhances model stability and supports fair comparison across different algorithms.

4.6 Chapter Summary

This chapter presented the implementation of the customer churn prediction framework. It began with exploratory data analysis, which provided insights into customer behaviour and highlighted key churn patterns related to contract type, service usage, and billing characteristics. The analysis also identified class imbalance and relationships among important numerical features, informing subsequent modelling decisions.

The chapter then described data transformation and feature engineering steps, including feature encoding, scaling, and the creation of interaction and aggregate features to enhance model learning. Finally, the chapter detailed the implementation and optimisation of multiple machine learning models, outlining training strategies, baseline and ensemble model development, hyperparameter tuning, and cross-validation procedures.

Overall, this chapter laid the groundwork for assessing model performance. The next chapter will present the experimental results and explore how effective the proposed models are in predicting customer churn.

Chapter 5: Results and Discussion

5.1 Performance Evaluation

5.1.1 Pre-Tuning Model Performance Analysis

This section describes the baseline performance of eight machine learning models trained on the balanced dataset, before any hyperparameter tuning was applied. Evaluation metrics such as accuracy, precision, recall, and F1-score were used to give a well-rounded assessment of each model's ability to predict. These results set a benchmark for comparing model performance and assessing the effects of future optimizations.

Table 1: ***Baseline Performance of Machine Learning Models***

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.7608	0.81	0.76	0.78
Random Forest	0.8318	0.83	0.83	0.83
LightGBM	0.8361	0.84	0.84	0.84
XGBoost	0.8332	0.83	0.83	0.83
Support Vector Machine	0.7871	0.82	0.79	0.80
K-Nearest Neighbours	0.7559	0.78	0.76	0.77
Gradient Boosting	0.8389	0.84	0.84	0.84
Decision Tree	0.7686	0.78	0.77	0.77

Observations

Boosting-based models (Gradient Boosting, LightGBM, and XGBoost) achieved the strongest baseline performance, indicating their superior ability to model complex, non-linear decision boundaries. Linear and distance-based models (Logistic Regression and KNN) exhibited comparatively lower performance, reflecting limited expressive capacity prior to optimisation. Random Forest delivered stable and competitive results, demonstrating the robustness of ensemble bagging methods even without tuning. Overall, the variation in baseline performance highlights the importance of hyperparameter optimisation to fully exploit each model's predictive potential.

5.1.2 Post-Tuning Model Accuracy Comparison

Following hyperparameter optimisation, all models were retrained and evaluated using three-fold cross-validation to obtain more reliable estimates of generalisation performance. The resulting accuracies reflect improvements achieved through systematic tuning of key parameters.

Table 2: ***Post-Tuning Cross-Validation Accuracy***

Model	Tuned CV Accuracy
XGBoost	0.8319
LightGBM	0.8296
Gradient Boosting	0.8286
Random Forest	0.8278
Decision Tree	0.7951
Support Vector Machine	0.7887
K-Nearest Neighbours	0.7713
Logistic Regression	0.7690

Observations

XGBoost achieved the highest tuned accuracy, closely followed by LightGBM and Gradient Boosting. Random Forest continued to demonstrate strong post-tuning performance, reinforcing the effectiveness of ensemble learning approaches.

Simpler models, including Logistic Regression, KNN, and Decision Tree, showed only marginal improvements, suggesting limited benefits from extensive tuning. Overall, optimisation improved model stability and narrowed performance gaps among top-performing algorithms.

5.2 Comparative Analysis of Model Performance

To further assess robustness and reliability, additional comparative analyses were conducted focusing on data splitting strategies, cross-validation stability, and tuning impact.

5.2.1 Impact of Train–Test Split Ratios

Model performance was evaluated using 80/20 and 50/50 train–test splits. Results indicated that models trained with the 80/20 split consistently achieved higher and more stable accuracy. The reduced training size in the 50/50 split limited learning capacity and increased variance. Consequently, the 80/20 split was selected as the optimal configuration for final evaluation and deployment.

--- Model Accuracy Comparison ---

	80/20 Split Accuracy	50/50 Split Accuracy
LightGBM	0.836054	0.829358
XGBoost	0.833215	0.823680
Random Forest	0.831796	0.823112
Logistic Regression	0.760823	0.765190

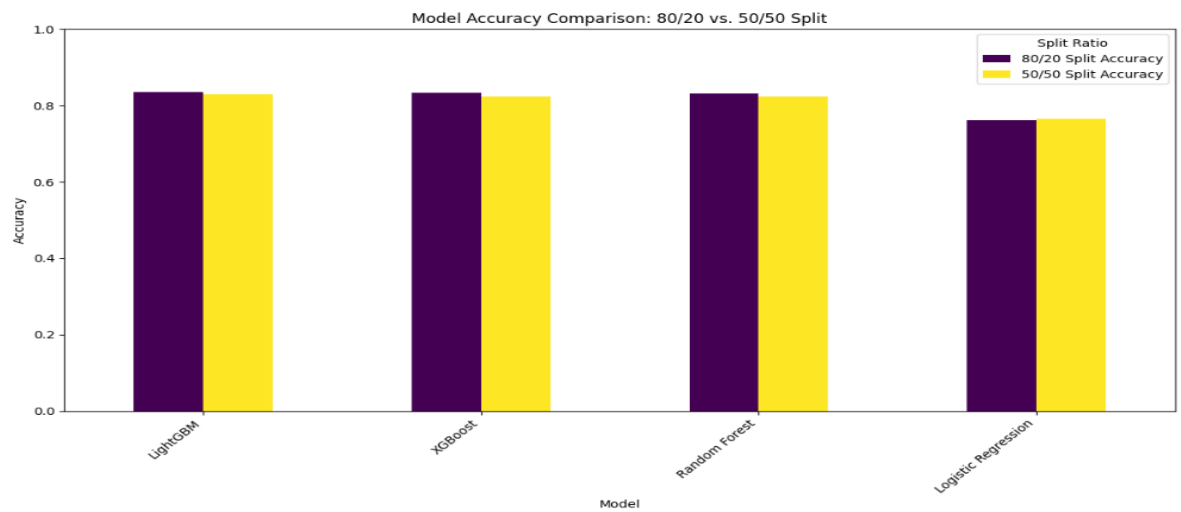

Figure 11: ***Model Accuracy Comparison Train–Test Splits***

Figure 11 compares model accuracy under two train–test split ratios (80/20 and 50/50). Across all models, the 80/20 split achieves higher accuracy, indicating improved learning capacity due to increased training data. This supports the selection of the 80/20 split for subsequent evaluation and final model deployment.

5.2.2 K-Fold Cross-Validation Stability

To reduce reliance on a single data split, K-Fold cross-validation was used. Models like XGBoost, LightGBM, Gradient Boosting, and Random Forest showed low variability across the folds, suggesting they have strong generalization ability. In contrast, Logistic Regression and KNN showed higher variance, suggesting sensitivity to data partitioning.

Model	Average Accuracy	Standard Deviation
Logistic Regression	0.7539	0.0108
Random Forest	0.8190	0.0068
LightGBM	0.8286	0.0031
XGBoost	0.8273	0.0035
Support Vector Machine	0.7838	0.0182
K-Nearest Neighbours	0.7514	0.0106
Gradient Boosting Classifier	0.8272	0.0060
Decision Tree Classifier	0.7667	0.0084

Table 3: ***5-Fold Cross-Validation Performance of Machine Learning Models***

From Table 3, the 5-fold cross-validation results indicate that ensemble-based models achieved superior and more consistent performance compared to simpler classifiers. LightGBM recorded the highest average accuracy (0.8286) with the lowest standard deviation (0.0031), closely followed by XGBoost and Gradient Boosting, demonstrating strong generalisation and stability across folds. Random Forest also performed reliably with moderate variance. In contrast, Logistic Regression, K-Nearest Neighbours, and Decision Tree classifiers showed lower average accuracy and higher variability, suggesting greater sensitivity to data partitioning. Overall, these results highlight the robustness of boosting-based ensemble methods for customer churn prediction.

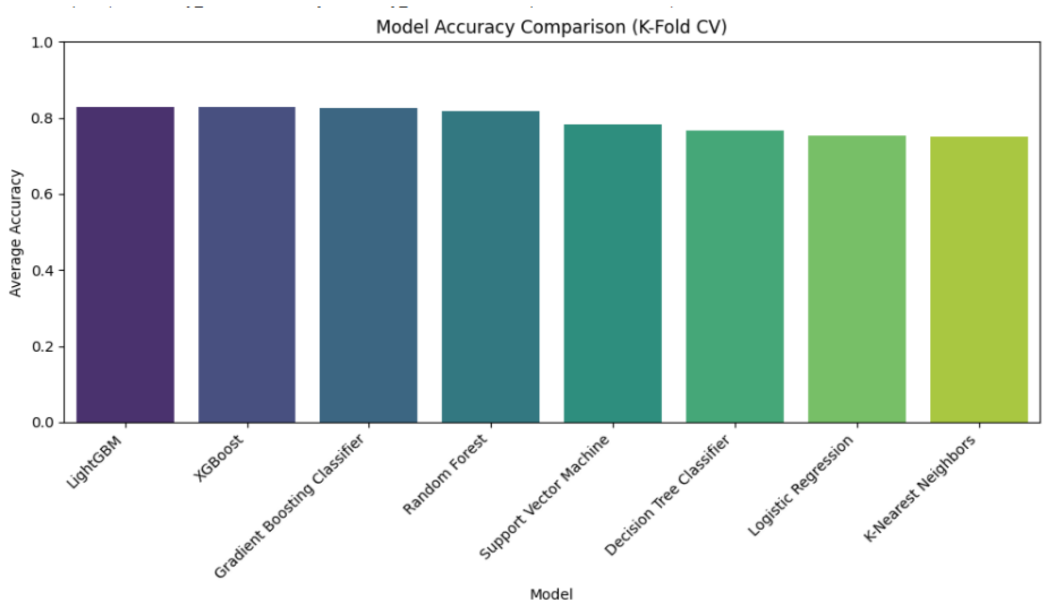

*Figure 12****: Average Model Accuracy Using 5-Fold Cross-Validation***

Figure 12 presents the average classification accuracy obtained from 5-fold cross-validation for all evaluated models. Ensemble-based methods, particularly LightGBM, XGBoost, and Gradient Boosting, achieved the highest average accuracy, demonstrating strong generalisation capability across folds. In contrast, simpler models such as Logistic Regression and K-Nearest Neighbours exhibited lower average accuracy, indicating reduced robustness to data partitioning. Overall, the results confirm that ensemble models provide more stable and reliable performance for customer churn prediction.

5.3 Effect of Hyperparameter Tuning

Hyperparameter tuning significantly enhanced model performance, particularly for tree-based ensemble models. Visual comparisons of pre- and post-tuning accuracy revealed clear improvements in predictive capability and stability. Simpler models benefited less from tuning, reinforcing the suitability of ensemble approaches for capturing complex churn patterns. Overall, tuning reduced performance variability and improved robustness across models.

*Table 4: **Hyperparameter Tuning Results Using RandomizedSearchCV (3-Fold CV)***

Model	Best CV Accuracy	Best Hyperparameters
Logistic Regression	0.7690	solver=liblinear, penalty=l2, C=10.0
Random Forest	0.8278	n_estimators=300, max_depth=10, max_features=log2, min_samples_split=5, min_samples_leaf=1
LightGBM	0.8296	n_estimators=100, learning_rate=0.05, num_leaves=31, max_depth=-1
XGBoost	0.8319	n_estimators=200, learning_rate=0.1, max_depth=6, subsample=1.0, colsample_bytree=0.8
Support Vector Machine	0.7887	kernel=rbf, C=5, gamma=scale
K-Nearest Neighbors	0.7713	n_neighbors=11, weights=distance, p=1
Gradient Boosting Classifier	0.8286	n_estimators=200, learning_rate=0.1, max_depth=3
Decision Tree Classifier	0.7951	max_depth=10, min_samples_split=5, min_samples_leaf=1

Note: The best-performing tuned model was XGBoost (0.8319), followed closely by LightGBM(0.8296) and Gradient Boosting (0.8286).

Table 4 summarises the results of hyperparameter tuning performed using RandomizedSearchCV with 3-fold cross-validation. The tuning process led to improved performance across all models, with ensemble methods showing the most notable gains. XGBoost achieved the highest tuned cross-validation accuracy, followed closely by LightGBM and Gradient Boosting, confirming the effectiveness of optimised ensemble learning for customer churn prediction.

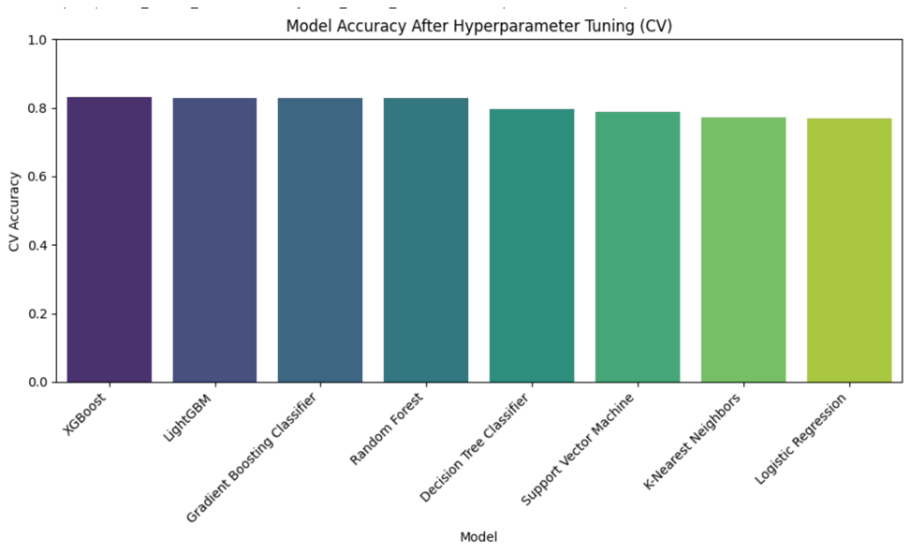

Figure 13: ***Model Accuracy After Hyperparameter Tuning (Cross-Validation)***

Figure 13 compares the cross-validated accuracy of all machine learning models after hyperparameter tuning. Boosting-based ensemble models, particularly XGBoost, LightGBM, and Gradient Boosting, achieved the highest accuracy, indicating that optimisation significantly enhanced their predictive capability. Random Forest also demonstrated strong performance, while simpler models such as Logistic Regression and K-Nearest Neighbours showed comparatively lower accuracy despite tuning. Overall, the figure confirms that hyperparameter optimisation benefits complex ensemble models more substantially than linear or distance-based classifiers.

5.4 Receiver Operating Characteristic (ROC) Analysis

ROC analysis using a one-vs-rest approach was conducted to evaluate classification performance beyond accuracy. XGBoost and LightGBM achieved the highest AUC values, indicating strong discriminative ability. Random Forest and Gradient Boosting also performed robustly, while Logistic Regression and KNN exhibited lower AUC scores, reflecting weaker class separation. These results further support the selection of boosting-based models for churn prediction.

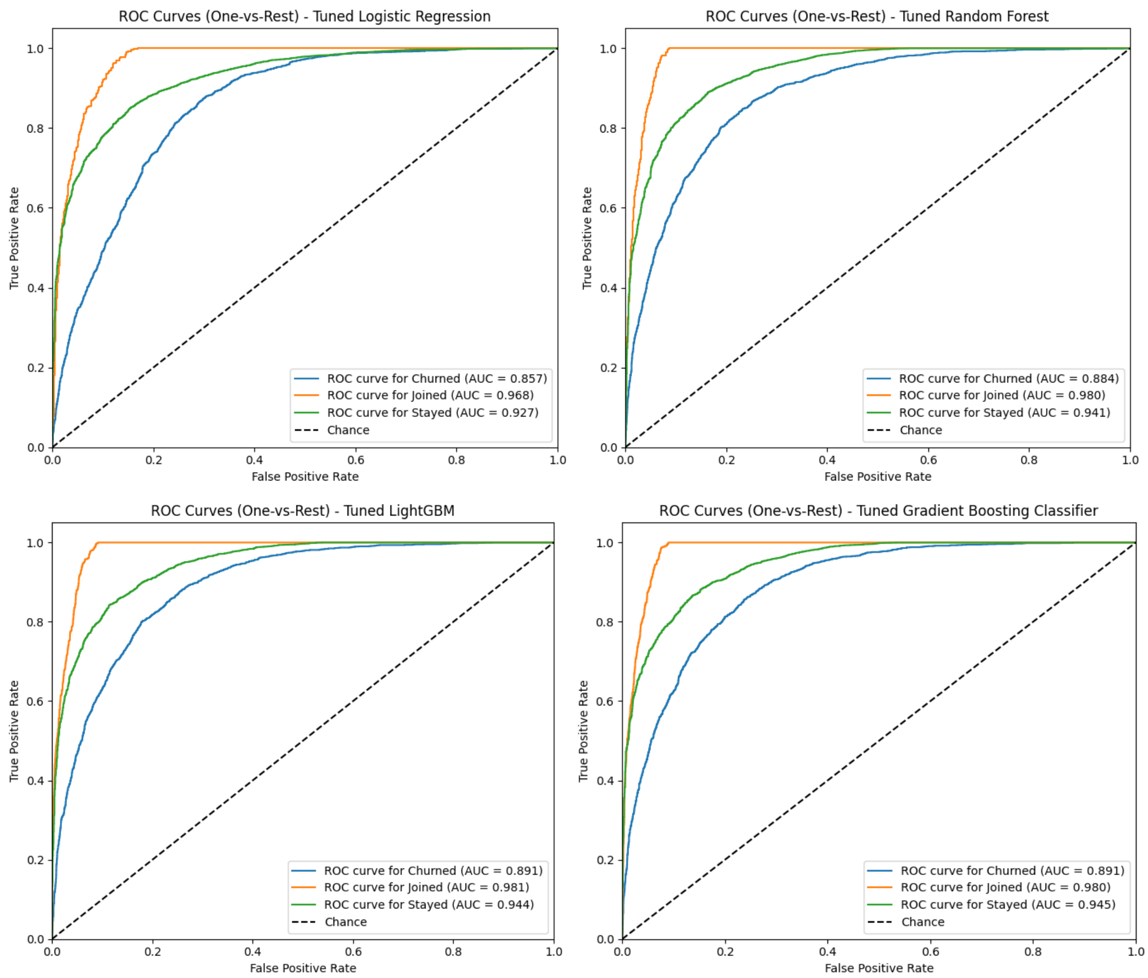

Figure 14: ***ROC Curves (One-vs-Rest) for Tuned Classification Models***

Figure 14 presents the one-vs-rest ROC curves for the tuned Logistic Regression, Random Forest, LightGBM, and Gradient Boosting Classifier models. For each model, separate ROC curves are shown for the *Churned, Joined,* and *Stayed* classes, along with the diagonal reference line representing random classification.

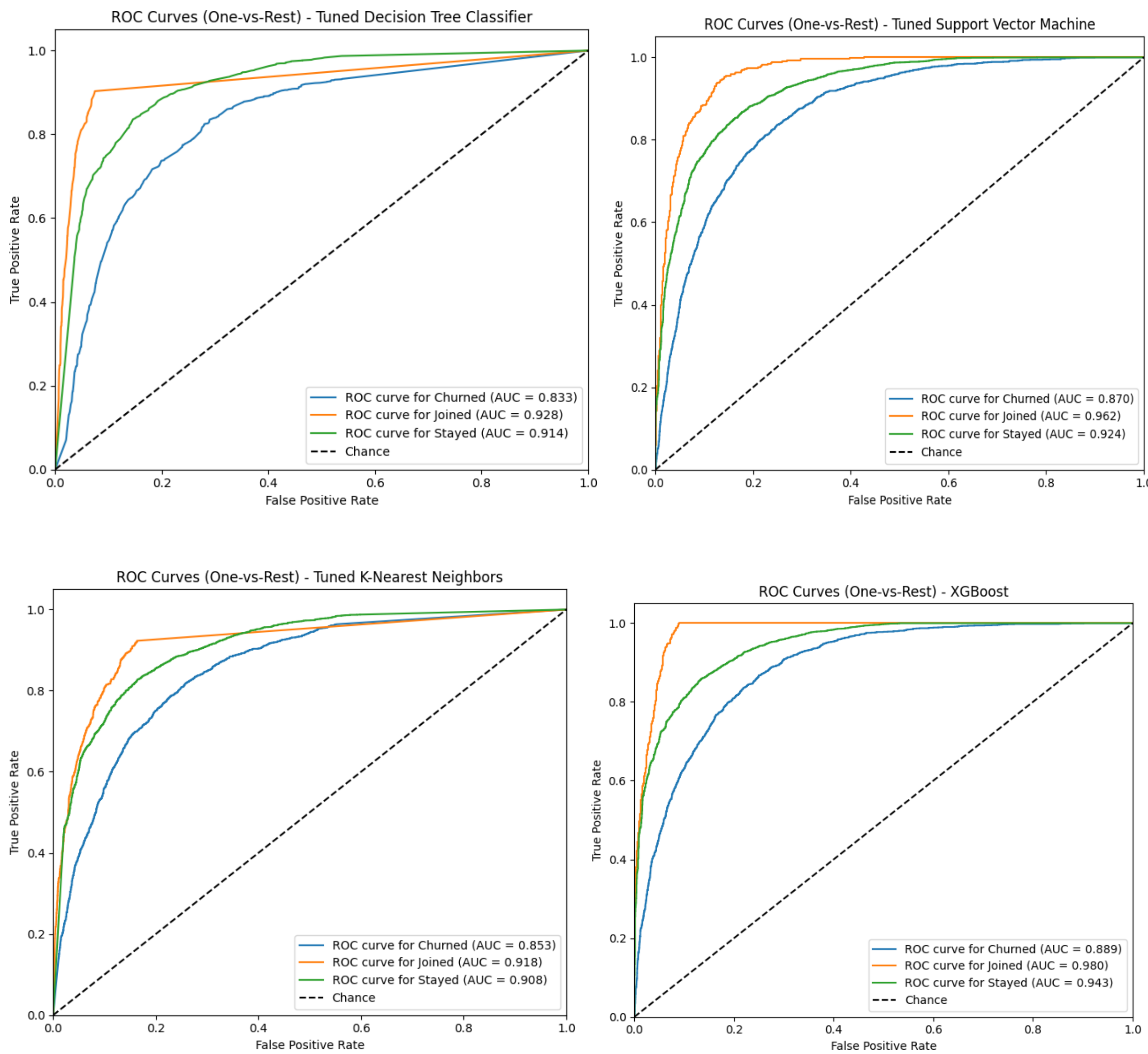

Figure 15: ***ROC Curves (One-vs-Rest) for Additional Tuned Models***

Figure 15 shows the one-vs-rest ROC curves for the remaining tuned classifiers: Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbours (KNN), and XGBoost. Separate ROC curves are presented for the *Churned, Joined,* and *Stayed* classes, with the diagonal line representing random classification performance.

Overall, the ROC curve analysis shows that ensemble-based models, particularly XGBoost, LightGBM, and Gradient Boosting, achieve higher AUC values across all customer classes, indicating stronger discriminative ability. The Churned class consistently exhibits lower AUC compared to the Joined and Stayed classes, highlighting the greater difficulty of churn prediction. Among all models, XGBoost demonstrates the most balanced and robust ROC performance, supporting its selection as the final model.

5.5 Confusion Matrix Analysis

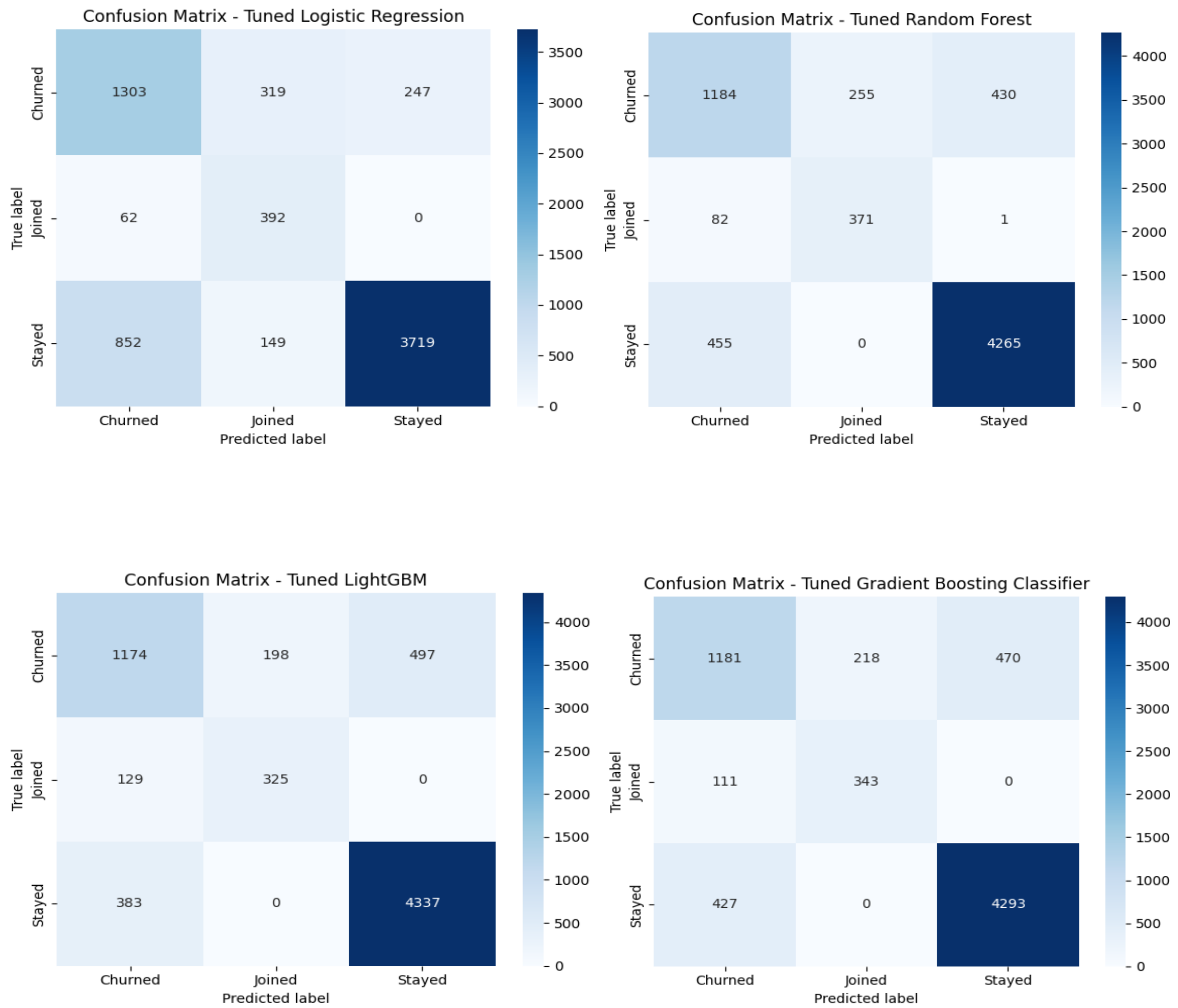

Figure 16: ***Confusion Matrices of Tuned Classification Models***

Figure 16 presents the confusion matrices for four tuned models: Logistic Regression, Random Forest, LightGBM, and Gradient Boosting Classifier. The matrices illustrate the distribution of correct and incorrect predictions across the three customer status classes (*Churned, Joined, Stayed*), providing detailed insight into each model's classification behaviour.

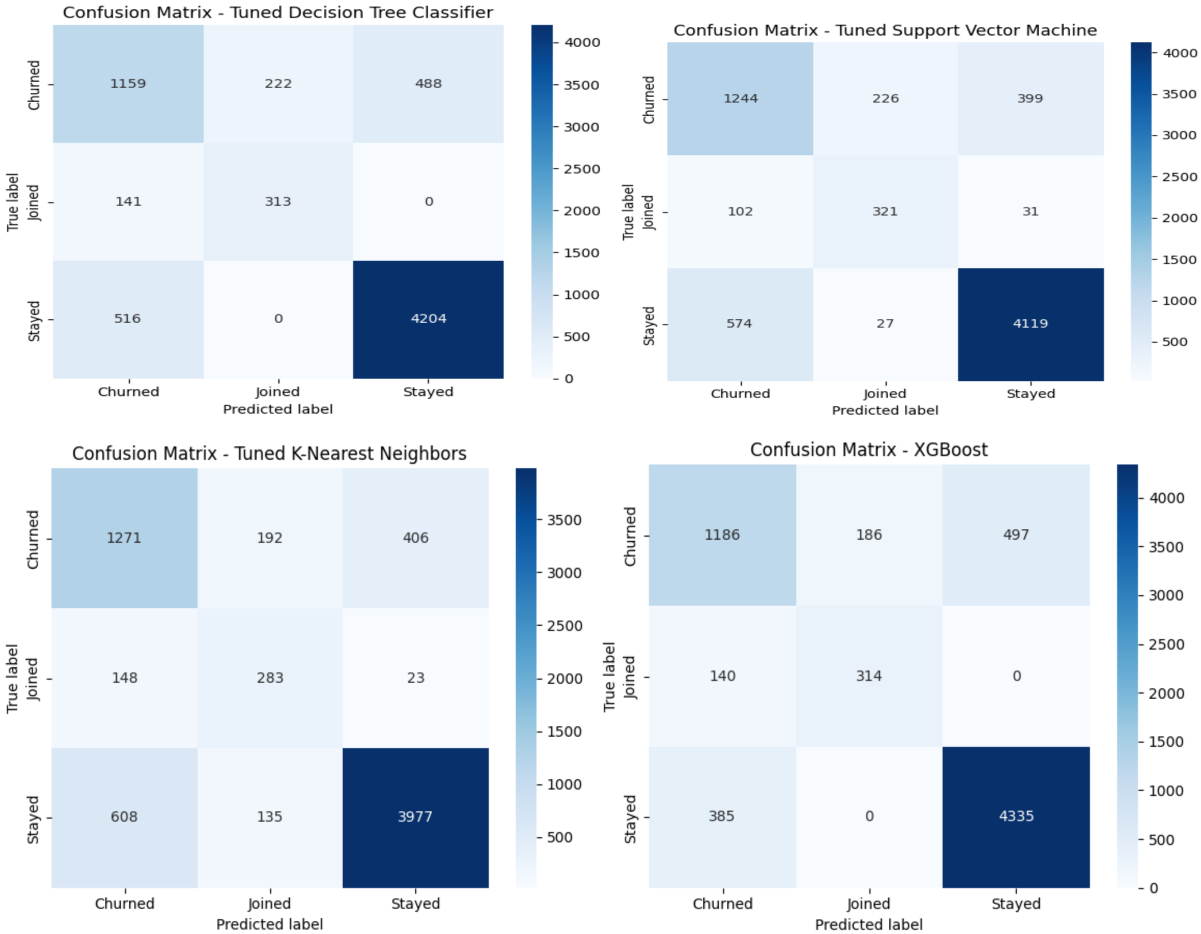

Figure 17: ***Confusion Matrices of Additional Tuned Models***

Figure 17 presents the confusion matrices for the remaining tuned classifiers: Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbours (KNN), and XGBoost. The matrices illustrate each model's ability to correctly classify customers into *Churned, Joined,* and *Stayed* categories.

5.6 Final Model Selection and Discussion

Based on combined evaluation across accuracy, cross-validation stability, ROC-AUC, and confusion matrix analysis, XGBoost was selected as the final model. The results demonstrate that:

- Boosting-based ensembles are highly effective for telecom churn prediction.
- Proper handling of class imbalance and hyperparameter optimisation significantly improves performance.

- Feature engineering and service-level attributes play a critical role in accurately identifying churn behaviour.

These findings align with existing churn prediction literature and validate the effectiveness of the proposed machine learning framework.

5.7 Model Explainability and Interpretation (SHAP and LIME)

While predictive performance metrics such as accuracy, ROC-AUC, and confusion matrices demonstrate how well the models perform, they do not explain *why* predictions are made. To address this limitation, Explainable AI (XAI) techniques were applied to the final XGBoost model to provide transparency and actionable insights into customer churn behaviour.

SHAP Global Feature Importance

SHAP analysis was applied to evaluate how each feature globally contributes to the churn prediction model. The SHAP summary plot indicates that Tenure in Months, Monthly Charges, Contract Type, Total Revenue, and Service-related features are the most influential variables driving model decisions. Lower tenure and higher monthly charges consistently increase churn probability, while longer contractual commitments reduce churn risk. These findings align with earlier descriptive analysis and reinforce the reliability of the predictive model**.**

SHAP Local Explanation

In addition to global explanations, SHAP was applied at the individual customer level to understand specific predictions. For example, customers with short tenure, month-to-month contracts, and limited-service subscriptions were assigned higher churn probabilities due to their strong positive SHAP contributions. Conversely, customers with long tenure and long-term contracts showed negative SHAP contributions toward churn, indicating retention likelihood. These local explanations demonstrate how the model combines multiple customer attributes to reach personalised churn predictions.

LIME Local Interpretation

LIME was used to further validate local explanations by approximating the XGBoost model behaviour around individual predictions. The LIME explanations highlighted similar influential features to SHAP, such as tenure, contract duration, and service engagement, confirming consistency between the two interpretability techniques. By providing human-readable explanations for individual predictions, LIME supports trust and usability, particularly for business stakeholders who require understandable decision rationales.

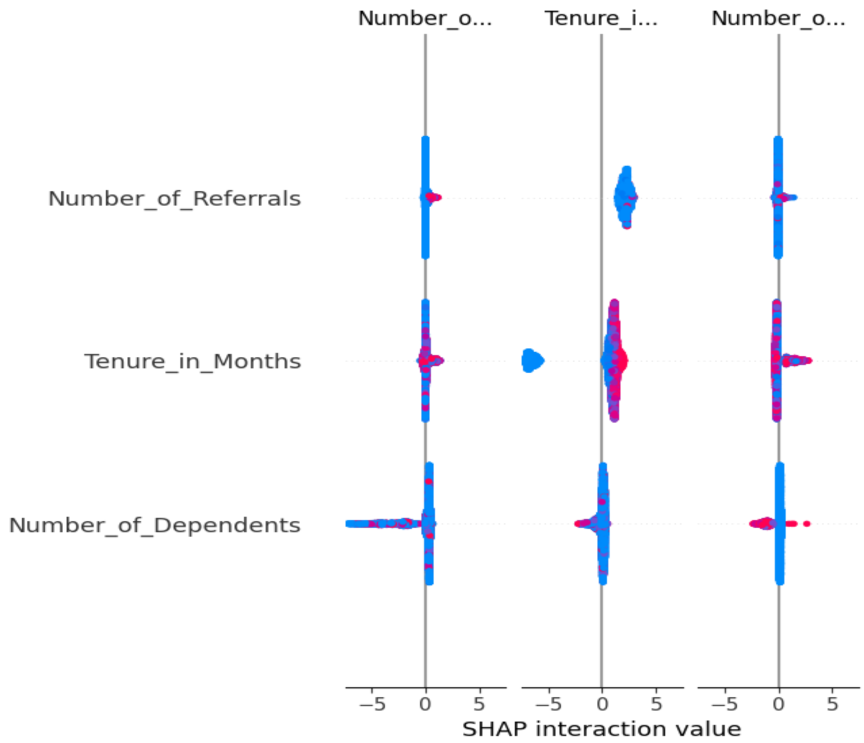

Figure 18: ***SHAP Interaction Plot for Key Customer Attributes***

Figure 18 illustrates the SHAP interaction effects between key customer attributes, including Tenure in Months, Number of Referrals, and Number of Dependents, and their combined influence on churn predictions. The plot shows how interactions between features contribute positively or negatively to the model's output rather than acting independently.

Higher tenure generally exhibits a stabilising effect, reducing churn probability, while interactions involving lower tenure amplify churn risk. Similarly, referral-related interactions indicate that customers with fewer referrals tend to contribute more positively to churn likelihood. These interaction patterns highlight the non-linear relationships captured by the XGBoost model and demonstrate its ability to model complex customer behaviour beyond simple feature importance.

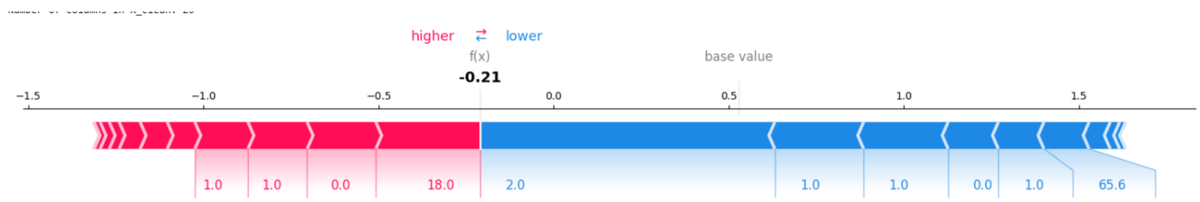

Figure 19: ***SHAP Local Explanation for an Individual Customer Prediction***

Figure 19 presents a SHAP force plot explaining an individual churn prediction generated by the XGBoost model. The plot illustrates how specific customer features contribute to pushing the prediction towards higher churn risk (red) or lower churn risk (blue) relative to the model's base value.

Features such as short tenure, low number of referrals, and limited service engagement contribute positively toward churn probability, while factors like higher monthly charges and certain service-related attributes reduce churn likelihood for this customer. This local explanation demonstrates how the model combines multiple feature effects to produce a final prediction, enhancing transparency and supporting customer-level decision-making.

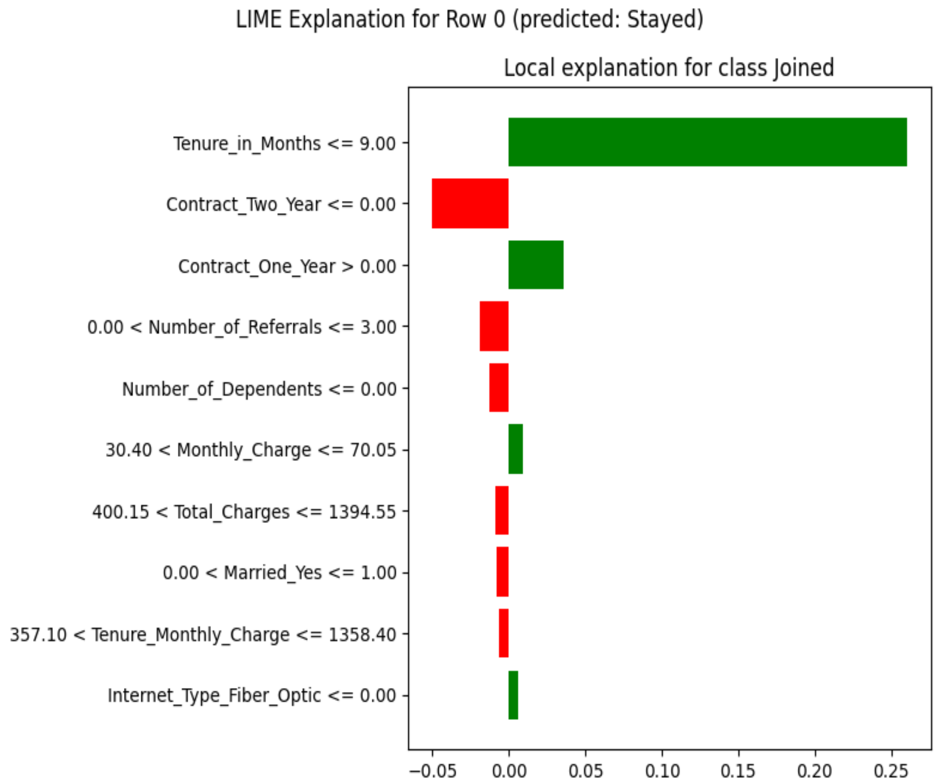

Figure 20: ***LIME Local Explanation for an Individual Customer Prediction***

Figure 20 presents a LIME explanation for an individual customer predicted as Stayed by the XGBoost model. The plot highlights the most influential features contributing to this prediction, with green bars indicating features that support the predicted class and red bars representing features that oppose it. The explanation shows that short tenure, contract duration, and referral-related attributes play a significant role in determining customer status. In this instance, tenure-related features and contract type contribute positively toward customer retention, while factors such as lower referrals and billing-related attributes exert a smaller opposing influence. This local interpretation complements SHAP results and demonstrates the model's ability to provide transparent, customer-level explanations for churn-related decisions.

Discussion

The integration of SHAP and LIME enhances the interpretability of the churn prediction framework, transforming the model from a black-box predictor into a transparent decision-support system. These explanations enable telecom providers to understand not only *which* customers are likely to churn, but also *why*, thereby supporting targeted retention strategies and informed business decisions. The explainability results strengthen the practical applicability and ethical reliability of the proposed system.

5.8 Model Deployment Using Streamlit

To demonstrate the practical applicability of the proposed churn prediction framework, the final tuned XGBoost model was deployed using the **Streamlit** framework as an interactive web-based application. The deployment enables real-time customer churn prediction by allowing users to input customer attributes through a graphical interface.

The application outputs the predicted customer status along with associated probability scores, supporting proactive decision-making. In addition, the deployment integrates model explainability components, enabling users to interpret predictions using SHAP and LIME visual explanations. This ensures transparency and trust in model decisions, particularly for non-technical stakeholders.

By deploying the model through Streamlit, the study extends beyond offline analysis and demonstrates how explainable machine learning models can be operationalised as decision-support tools in real-world telecom environments.

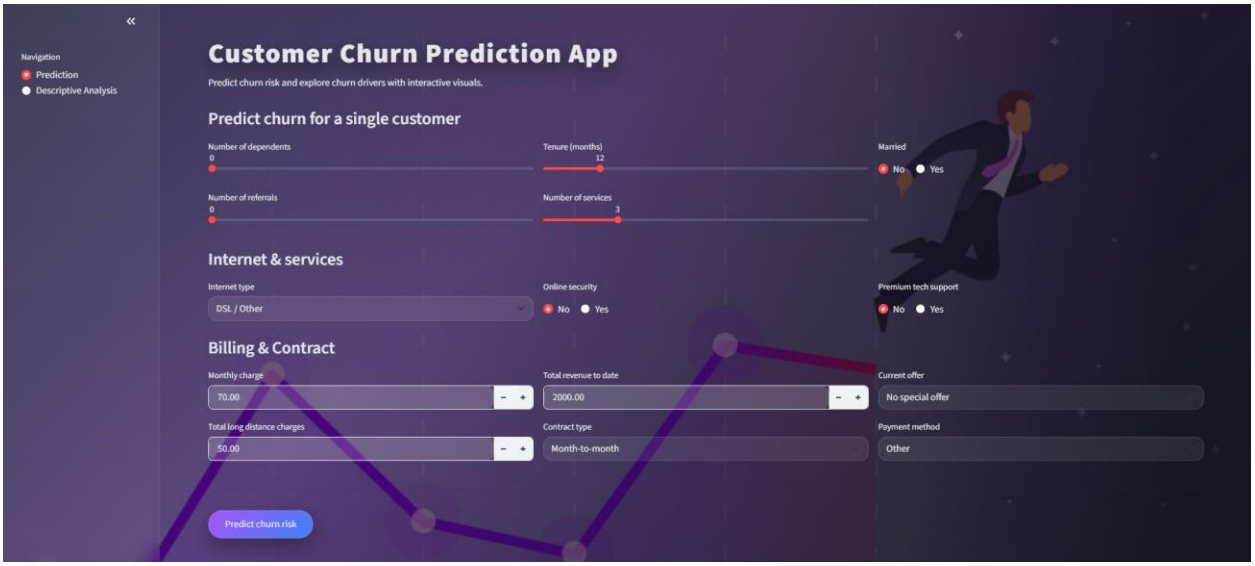

*Figure 21**: Streamlit-Based Customer Churn Prediction Dashboard***

Figure 21 presents the deployed Streamlit dashboard developed for customer churn prediction. The interface allows users to input customer demographic, service usage, and billing-related attributes through an interactive form. Based on the provided inputs, the application generates real-time churn predictions using the final tuned XGBoost model. The dashboard supports intuitive exploration of churn risk while maintaining transparency by enabling integration with explainable AI outputs. This deployment demonstrates the practical applicability of the proposed machine learning framework and its suitability for real-world telecom decision-support systems.

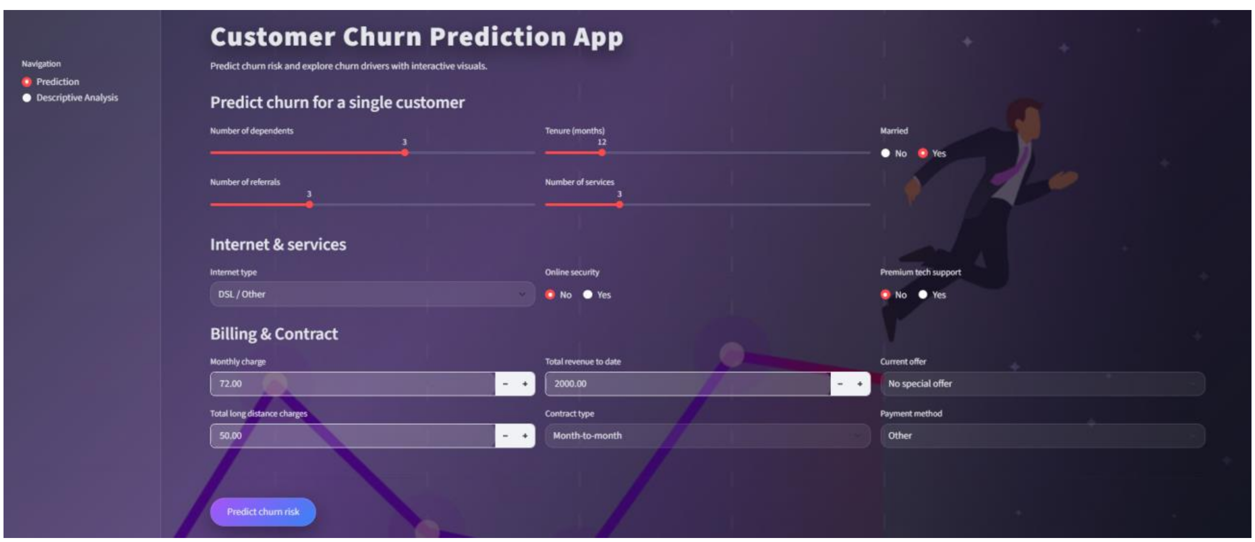

*Figure 22**: Interactive Customer Input Interface for Churn Prediction (Streamlit Application)***

Figure Y illustrates the testing phase of the deployed Streamlit-based customer churn prediction application. The dashboard allows users to modify customer-specific attributes, including demographic details, service usage, contract information, and billing characteristics, using interactive sliders, dropdowns, and input fields. In this example, the customer profile has been adjusted to represent a realistic telecom user scenario prior to prediction. Upon clicking the *Predict churn risk* button, the trained and tuned XGBoost model generates a churn probability and classification outcome in real time. This interactive testing capability demonstrates how the proposed system can support scenario analysis and proactive churn risk assessment in practical business settings.

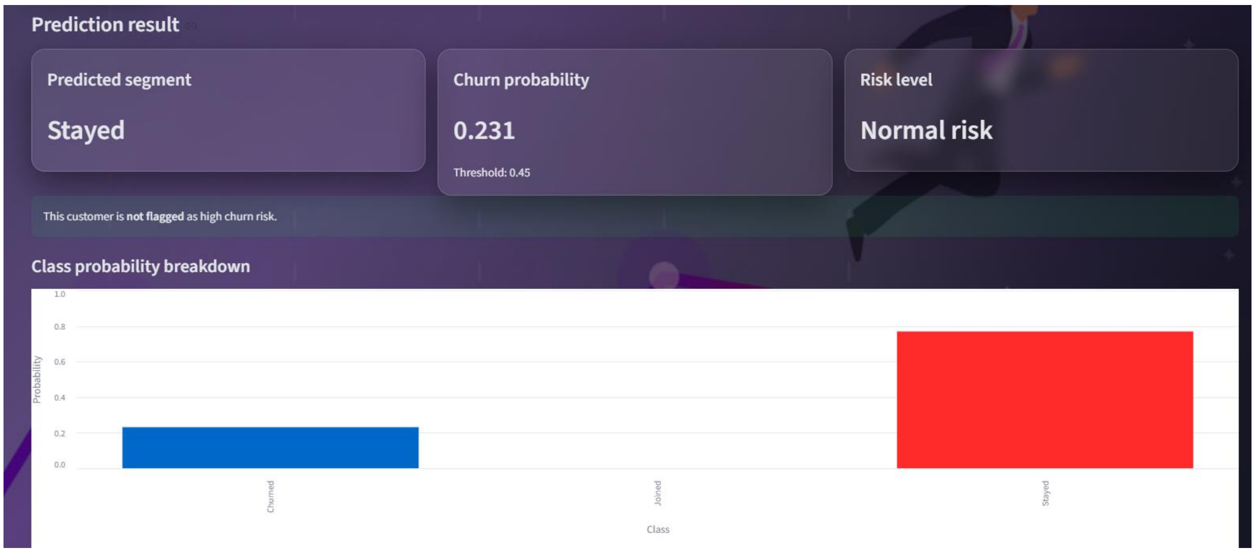

Figure 23: ***Customer Churn Prediction Output and Risk Interpretation (Streamlit Application)***

Figure 23 presents the prediction results generated by the deployed Streamlit-based churn prediction system after submitting customer input data. The dashboard displays the predicted customer segment (*Stayed*), the estimated churn probability (0.231), and the corresponding risk level (*Normal risk*), based on a predefined decision threshold of 0.45. A class probability breakdown is also visualised, showing the model's confidence across the Churned, Joined, and Stayed classes. This output demonstrates how the trained XGBoost model translates complex feature interactions into interpretable risk scores, enabling stakeholders to quickly assess churn likelihood and prioritise customer retention actions.

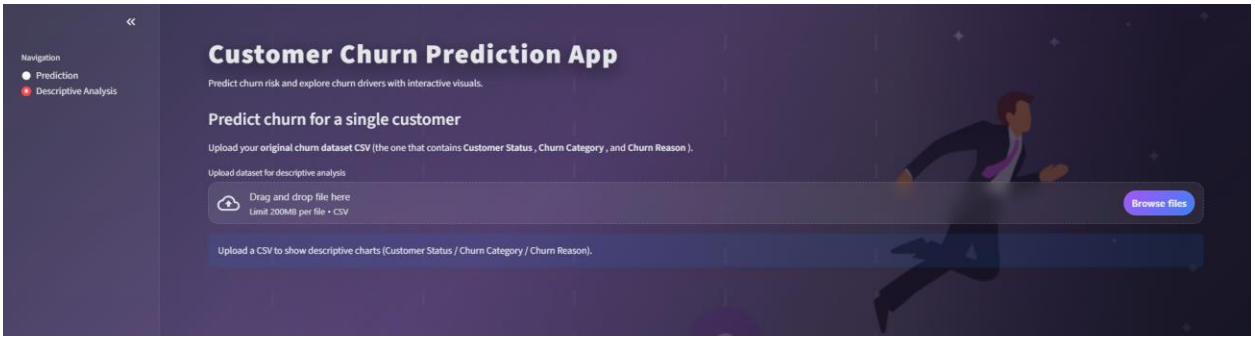

Figure 24: ***Descriptive Analysis Interface of the Streamlit-Based Churn Prediction Application***

Figure 24 shows the *Descriptive Analysis* interface of the deployed Streamlit application. This module allows users to upload the original telecom churn dataset containing Customer Status, Churn Category, and Churn Reason variables. Once the dataset is uploaded, the application dynamically generates descriptive visualisations, including class distributions and churn reason summaries, enabling exploratory analysis directly within the web interface. This functionality bridges offline exploratory data analysis and real-time interactive analytics, allowing stakeholders to examine churn patterns, dominant churn categories, and key churn drivers without requiring technical expertise or direct access to the underlying code.

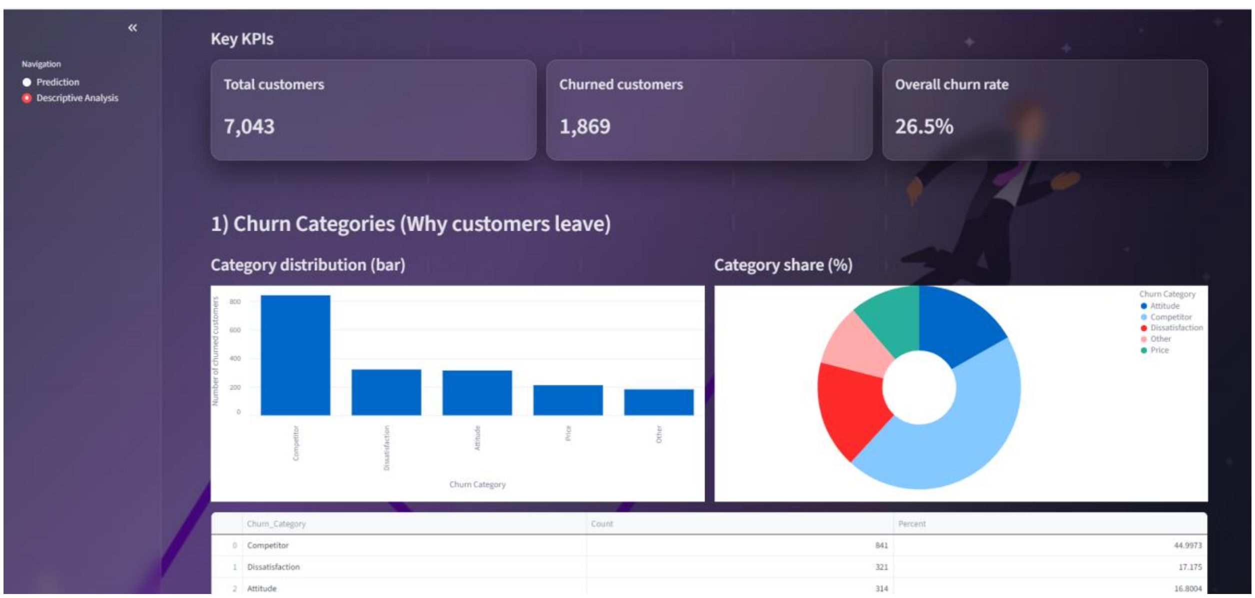

Figure 25: ***Descriptive Analysis Dashboard for the Full Telecom Churn Dataset***

Figure 25 presents the descriptive analysis dashboard generated within the Streamlit-based customer churn prediction application after uploading the complete telecom dataset. The dashboard displays key performance indicators (KPIs), including the total number of customers (7,043), the number of churned customers (1,869), and the overall churn rate (26.5%). In addition, customer status distribution is visualised using both pie and bar charts, showing the proportion of customers who have stayed, churned, or joined the service. This interactive dashboard enables users to explore dataset-level churn patterns in real time, providing an intuitive summary of churn prevalence and class imbalance before predictive modelling and explanation.

CHAPTER 6: CONCLUSION AND FUTURE WORK

6.1 Conclusion

This project focused on developing a machine learning–based solution to predict customer churn in the telecom sector. Customer churn remains a critical challenge for telecom providers due to its direct impact on revenue, customer lifetime value, and operational costs. The primary objective of this study was to analyse customer behaviour and service usage patterns, and to build predictive models capable of identifying customers at high risk of churn.

The project followed a structured machine learning pipeline beginning with data preprocessing, exploratory data analysis, and feature engineering. Missing values were handled appropriately, categorical variables were encoded, and class imbalance was addressed using the Synthetic Minority Over-sampling Technique (SMOTE). Exploratory analysis revealed that contract type, tenure, internet service characteristics, monthly charges, and value-added services such as online security and technical support play a significant role in influencing churn behaviour.

Multiple machine learning algorithms were evaluated, including Logistic Regression, Random Forest, Gradient Boosting, LightGBM, XGBoost, Support Vector Machine, K-Nearest Neighbors, and Decision Tree classifiers. Baseline model performance indicated that ensemble and boosting-based models consistently outperformed simpler algorithms. Following hyperparameter tuning and cross-validation, XGBoost emerged as the most effective model, demonstrating strong generalisation capability, high accuracy, and superior discrimination between churned and non-churned customers.

The results confirm that machine learning models, particularly boosting-based approaches, are highly effective for churn prediction when supported by proper preprocessing, feature selection, and model optimisation. The findings of this study provide actionable insights that telecom providers can leverage to design targeted retention strategies, optimise service offerings, and reduce customer attrition.

6.2 Business Implications

The results of this project have important practical implications for telecom providers. By accurately pinpointing customers who are likely to churn, companies can take proactive steps to retain them, such as offering personalized deals, contract incentives, or service upgrades. The analysis highlights the importance of bundled services, long-term contracts, and customer engagement in reducing churn. Additionally, understanding the influence of payment methods and internet service types enables providers to address dissatisfaction points more effectively.

The predictive framework developed in this project can be integrated into customer relationship management (CRM) systems to support real-time decision-making and improve customer satisfaction.

6.3 Limitations of the Study

Despite the promising results, this study has several limitations. Firstly, the dataset represents a snapshot of customer behaviour and does not capture temporal changes or evolving customer preferences over time. Secondly, the model relies solely on structured data and does not incorporate unstructured sources such as customer feedback, call transcripts, or complaint logs, which could provide deeper behavioural insights. Additionally, while SMOTE effectively addressed class imbalance, synthetic sampling may introduce noise that could affect real-world deployment.

6.4 Future Work

There are several opportunities to extend and enhance this research in future studies:

1. **Time-Series and Sequential Modelling** Future work could incorporate time-series analysis or recurrent neural networks (RNNs) to model customer behaviour overtime, enabling early churn detection.

2. **Incorporation of Unstructured Data** Integrating text data from customer support interactions, reviews, or surveys using Natural Language Processing (NLP) techniques could improve prediction accuracy.

3. **Advanced Deep Learning Models** Exploring deep learning approaches such as neural networks or attention-based models may further enhance predictive performance, particularly with larger datasets.

4. **Cost-Sensitive Learning** Future models could incorporate cost-sensitive learning to prioritise high-value customers and minimise financial loss due to churn.

5. **Real-Time Deployment and Monitoring** Deploying the model as a real-time API or dashboard would enable continuous monitoring and proactive churn prevention.

6. **Explainable AI (XAI)** Implementing explainability techniques such as SHAP or LIME would improve model transparency and help business stakeholders understand churn-driving factors.

6.5 Final Remarks

This project demonstrates the effectiveness of machine learning techniques in addressing real-world business challenges such as customer churn prediction. By combining data-driven insights with advanced modelling approaches, telecom providers can enhance customer retention and improve long-term profitability. The methodologies and findings presented in this study provide a solid foundation for future research and practical deployment in customer analytics systems.

References

D. Y. C. Wang et al., “Explainability of highly associated fuzzy churn patterns,” *arXiv*, 2024.

J. Sana et al., “Privacy-preserving customer churn prediction model,” *arXiv*, 2024.

M. A. Shaikhsurab and P. Magadum, “Adaptive ensemble learning for telecom churn prediction,” *arXiv*, 2024.

“Explaining customer churn prediction in telecom industry,” ScienceDirect, 2024.

“Customer churn prediction model using explainable …” *arXiv* (J. Maan), 2023.

S. Pandikumar & M. Shoukath B H, “Explainable AI approach for telecom churn,” IJSREM, 2025.

“Comparative analysis of XAI techniques on telecom churn,” *ResearchGate*, 2024.

D. Asif and M. S. Arif, “A data-driven approach with explainable artificial intelligence for customer churn prediction in the telecommunications industry,” *Results in Engineering*, vol. 26, 2025.

S. Liu, “Literature review on customer churn prediction in telecom industry,” *CONF-APMM Symposium 2025*, 2025.

J. Yuan, “Telecom customer churn prediction with explainable machine learning,” *ACM Digital Library*, 2025.

D. Kumari et al., “Telecom customer churn forecasting using machine learning: A data-driven predictive framework,” Metallurgical and Materials Engineering Research, 2025.

P. Boozary, “Enhancing customer retention with machine learning,” *ScienceDirect*, 2025.

“An explainable AI approach for telecom churn prediction and retention strategy design using XGBoost and SHAP,” IJSREM, 2025.

V. Chang et al., “Prediction of customer churn behavior in ... using XAI techniques,” *Algorithms*, vol. 17, no. 6, 2024.

M. Imani, “Customer churn prediction: Systematic review,” *MDPI*, 2025.

C. Özkurt, “The role of explainable AI (XAI) in customer churn analysis,” *ITEB Scientific Journal*, 2025.

S. Kammili, *Customer segmentation and churn prediction in telecommunication using machine learning*, MSc Thesis, NCI, 2025.

Verbeke, W., Van den Poel, D., & Lescot, J. (2012). *Predicting customer churn in the telecommunications industry using a novel approach based on variable selection and ensemble learning.* Journal of Machine Learning Research, 13, 1627-1651.

Powers, D. M. W. (2011). *Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation.* Journal of Machine Learning Technologies, 2(1), 37-63.

Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable* (2nd ed.). Springer.

S. Bhushan, *Enhancing customer churn prediction in the telecom sector*, MSc Thesis, NCI, 2025.

T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2016)*, San Francisco, CA, USA, 2016, pp. 785-794.

A. Prashanthan, “A deployable machine learning pipeline with explainable AI for churn,” *FAITH Journal*, 2025.

Y. Wang, “Advancements of customer churn in the telecommunications ...” *ECAI 2024*, 2025.

A. AbdelAziz, “A comprehensive evaluation of machine learning and deep learning for churn prediction,” *Information*, vol. 16, 2025.

“Explainable machine learning models applied to predicting customer churn for e-commerce,” *ResearchGate*, Dec. 2024.

A. Verbeke et al., “New insights into churn prediction in the telecommunication sector,” *Eur. J. Oper. Res.*, 2012.

H. He and E. A. Garcia, “Learning from imbalanced data,” *IEEE T. Knowl. Data Eng.*, vol. 21, no. 9, 2009.

N. V. Chawla et al., “SMOTE: Synthetic minority over-sampling technique,” *J. Artif. Intell. Res.*, vol. 16, 2002.

F. Doshi-Velez and B. Kim, “Towards a rigorous science of interpretable machine learning,” arXiv preprint, 2017.

R. Guidotti et al., “A survey of methods for explaining black box models,” *ACM Comput. Surv.*, vol. 51, no. 5, 2018.

A. B. Arrieta et al., “Explainable AI (XAI): Concepts, opportunities and challenges,” *Information Fusion*, vol. 58, 2020.

G. Ke et al., “LightGBM: A highly efficient gradient boosting decision tree,” *NeurIPS*, 2017.

N. M. AbdelAziz et al., “Comprehensive evaluation of advanced ML & DL models for churn,” *MDPI*, 2025.

Lundberg, S. M., & Lee, S. I. (2017). *A unified approach to interpreting model predictions.* In Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS 2017), 4765-4774.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). *"Why should I trust you?" Explaining the predictions of any classifier.* In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2016), 1135-1144.