Phase 3 Project Presentation

John Kul

Overview

The project intends to buffer the customer loss of a telecommunication company. The project is done based on different customer-related attributes to construct a machine learning model that will classify a customer as "churn" or "not churn".

The major stakeholder of the project is the Customer Retention Team of a telecommunication company. It will seek to reduce customer loss by highlighting the at-risk customers and giving indications for the application of retention initiatives.

Business Problem

Problem Statement: Customer churn is one of the biggest dilemmas in the field of telecommunication, through which revenue might get a chance to be lost, and customer acquisition costs go up. In reference, the business problem will be to develop a predictive model that accurately identifies customers who are at risk of churning. This may provide an opportunity for the retention team to act proactively by reaching out to customers with some form of incentives to reduce churning rates.

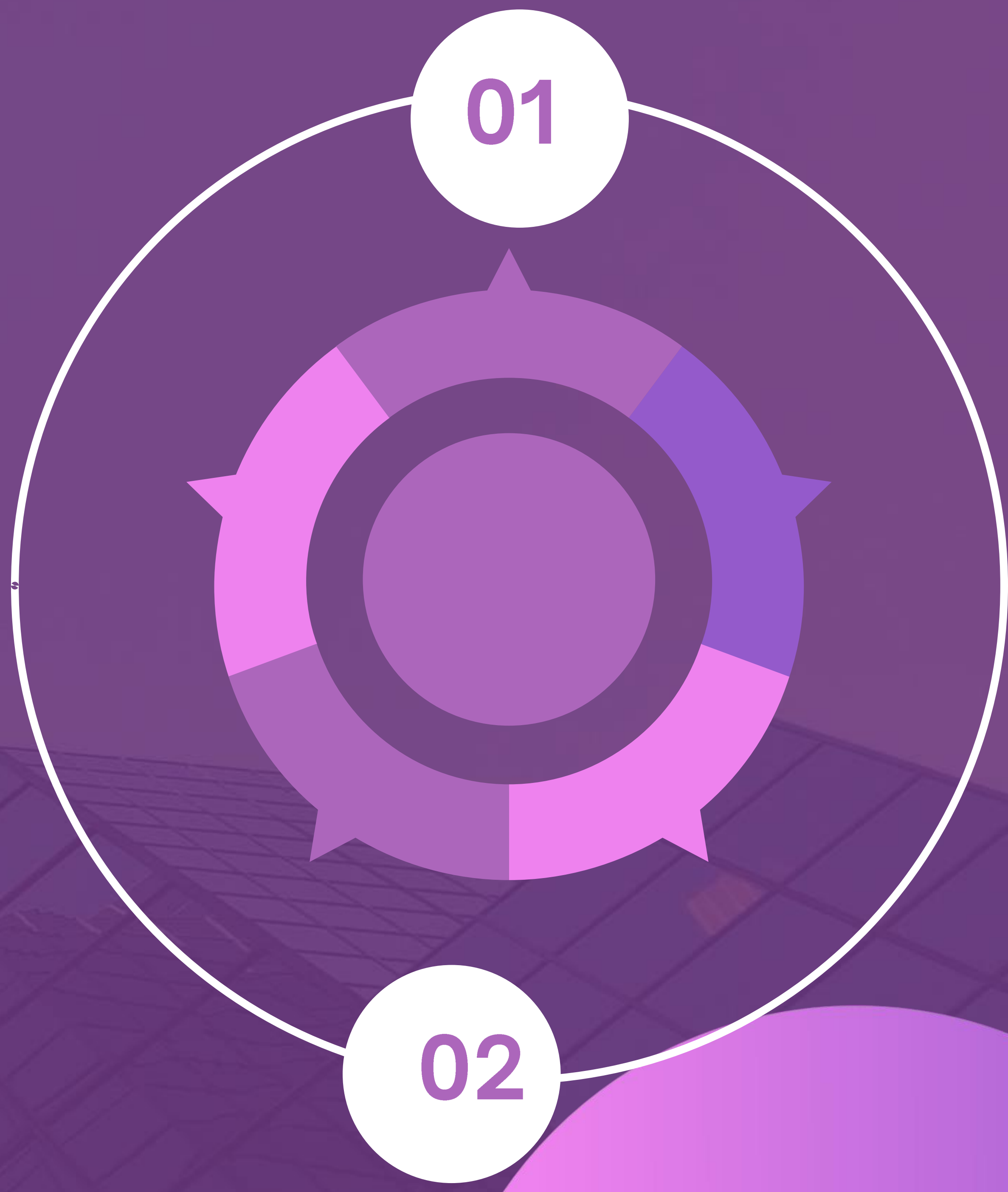
Data

The dataset includes the following information for customers: how long the customer has had an account, their usage patterns, meaning total minutes and charges for day, evening, and night, whether international and voicemail plans are on, and interactions with customer service. The target variable is churn, which describes whether the customer has left or not.

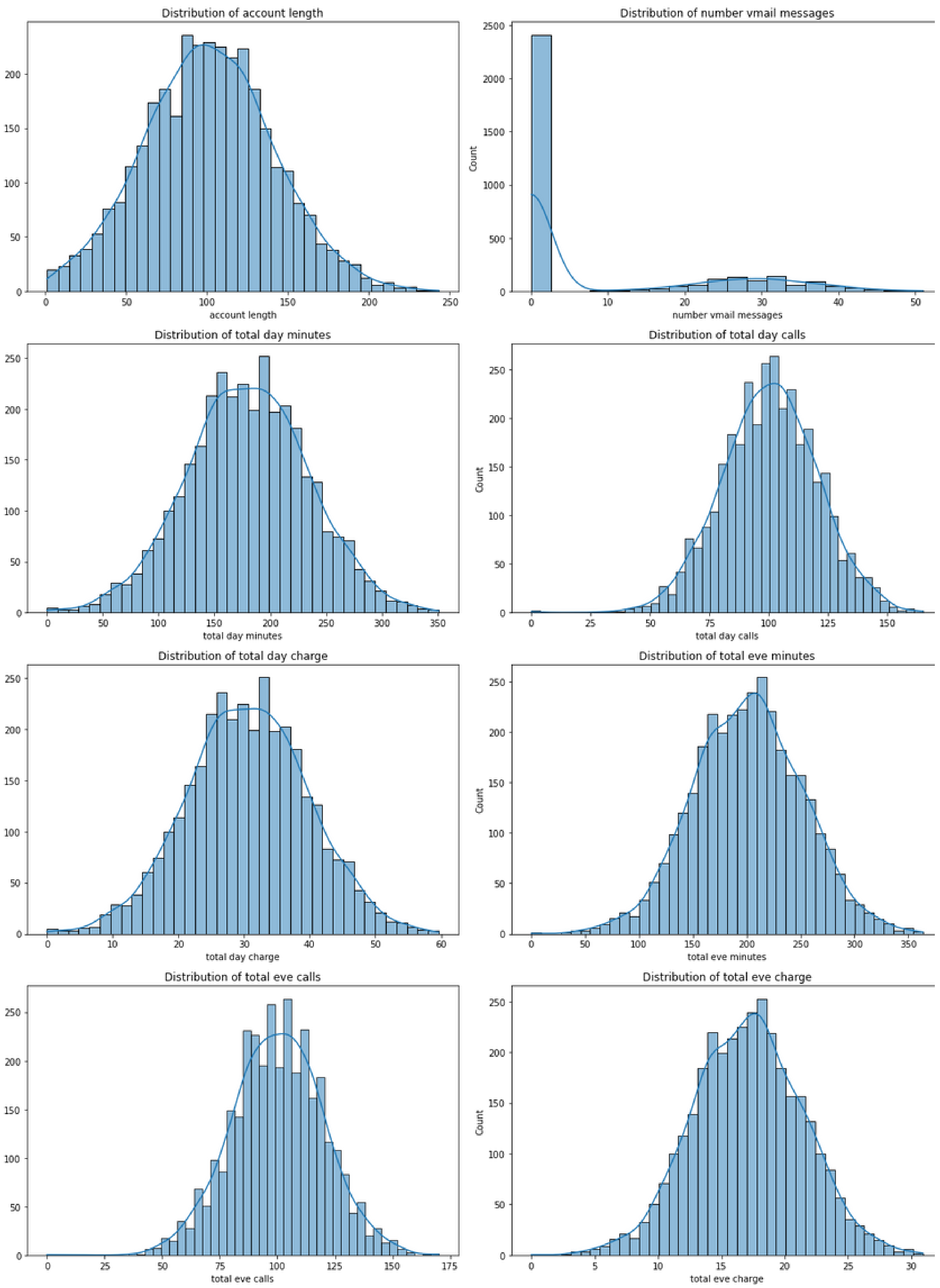
Classification Task Overview

Target Variable: It is a binary categorical variable because it contains either True, representing the customers who have churned, or False for those who have not.

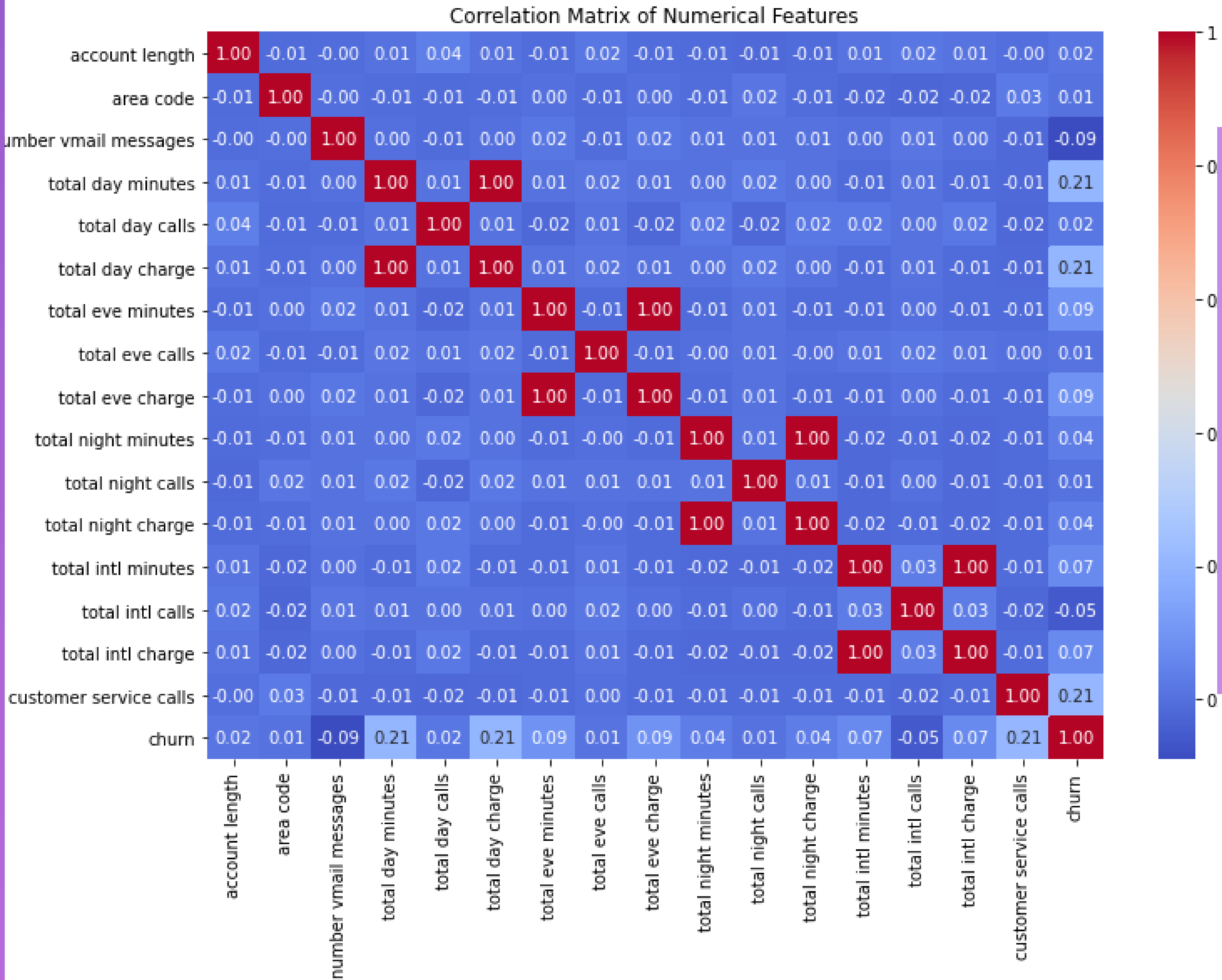
Classification Problem: Prediction for whether a customer will churn TRUE or not FALSE. Since the target variable is of categories-a customer has either churned or not churned-it falls under a classification problem, rather than a regression problem.

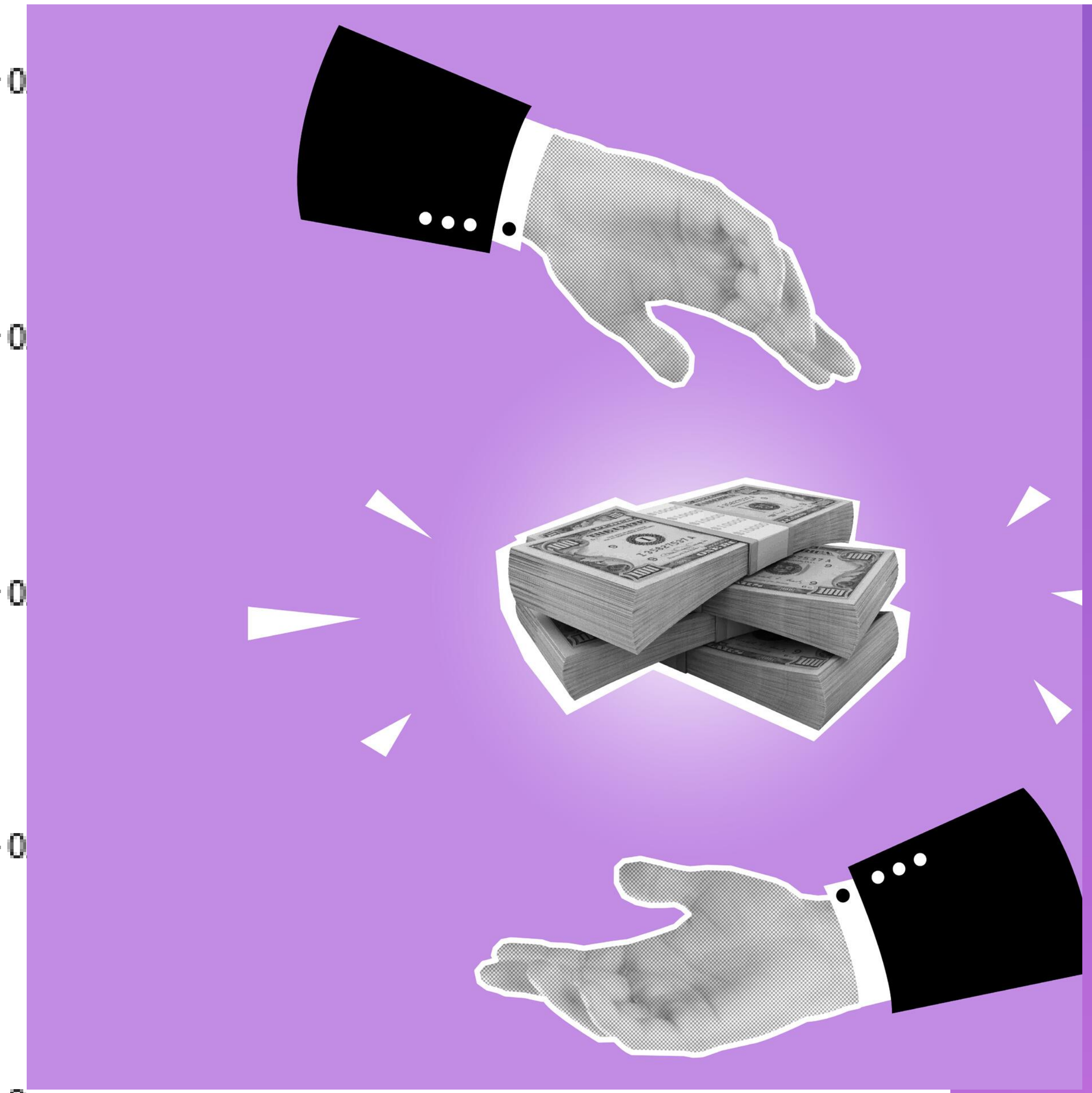


Ploting the Distribution of the numerical features of the churn dataset.



correlation of the numerical features





Interpretation

Distributions: These numeric features such as total day minutes, total eve minutes, and total night minutes are a bit right-skewed, which means most of the customers use less than the maximum available minutes. This includes features like number vmail messages and customer service calls which all have a high number of zeroes, probably indicating that a large number of customers either do not ever use voicemail or very seldom call customer service.

Analysis of Correlation: The correlation matrix shows strong features of inter-feature dependencies involving the features total day minutes, total day charge, total eve minutes, and total eve charge. This makes a lot of sense because these charges are actually derived from the respective minutes. Most of the features are weakly correlated with the target variable of churn, hence suggesting more complex modeling to capture this relationship.

FAUGET

MODELLING

Baseline Logistic Regression Model Performance

Accuracy: 0.86

	precision	recall	f1-score	support
False	0.86	1.00	0.92	857
True	0.00	0.00	0.00	143
accuracy			0.86	1000
macro avg	0.43	0.50	0.46	1000
weighted avg	0.73	0.86	0.79	1000

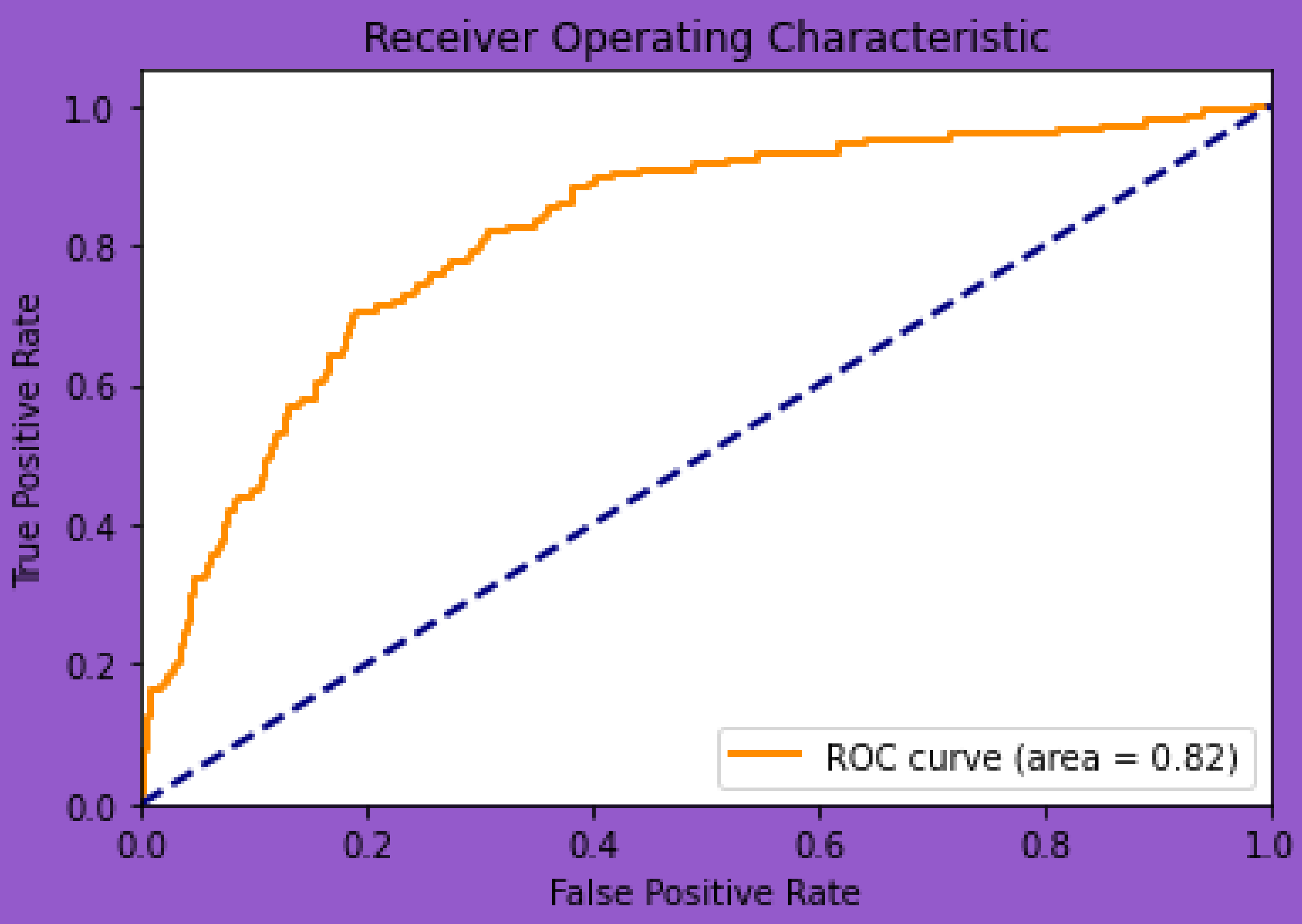
Tuned Logistic Regression Model Performance

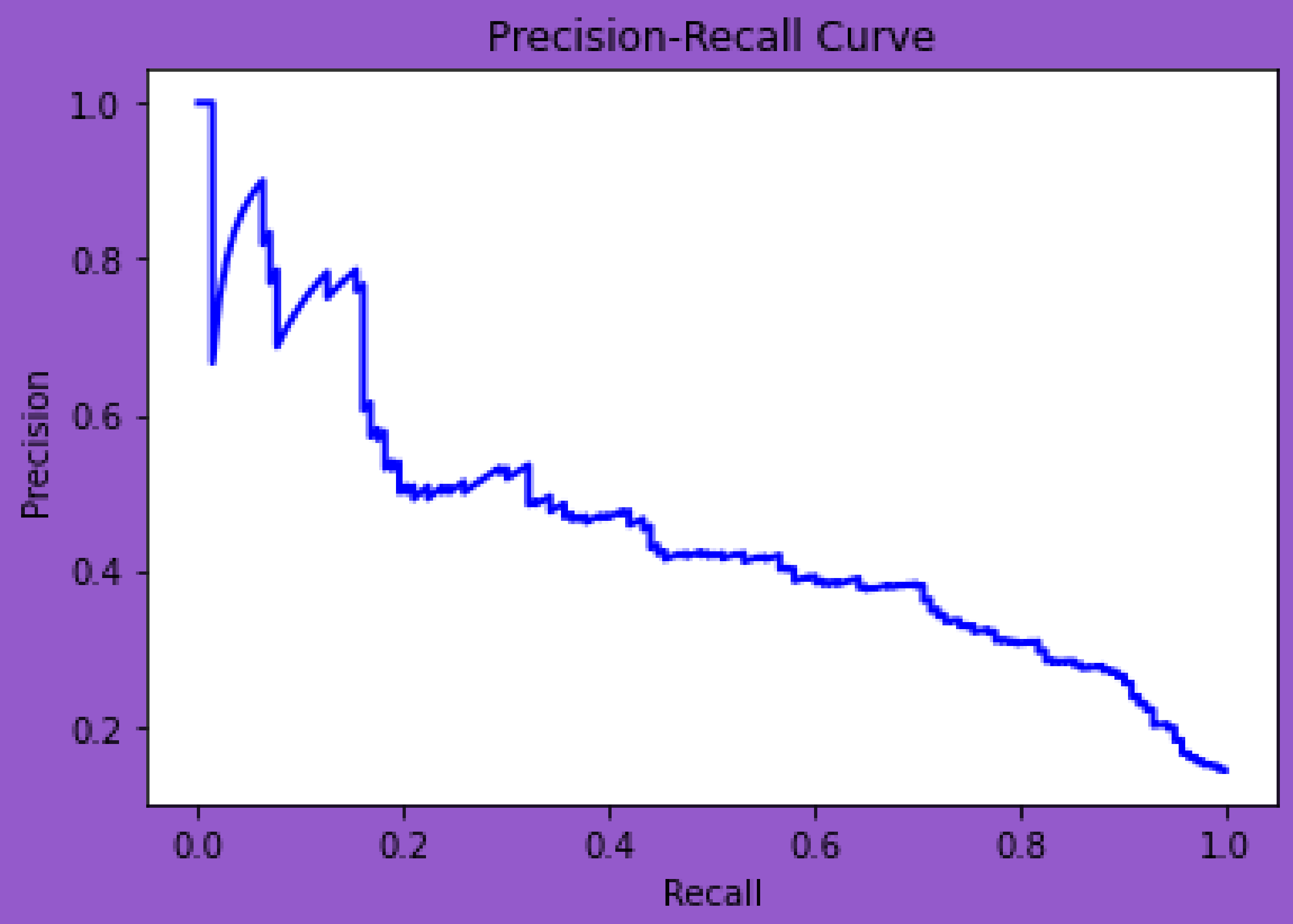
Accuracy: 0.86

	precision	recall	f1-score	support
False	0.86	1.00	0.92	857
True	0.00	0.00	0.00	143
accuracy			0.86	1000
macro avg	0.43	0.50	0.46	1000
weighted avg	0.73	0.86	0.79	1000

The baseline logistic regression model performs well in terms of accuracy (0.86) and handles the majority class (False) effectively, with a precision of 0.86 and recall of 1.00. However, it completely fails to identify the minority class (True), with both precision and recall at 0.00, resulting in an F1-score of 0.00 for this class. This highlights the model's severe bias towards the majority class and its inability to manage class imbalance, making it unsuitable for tasks where the minority class is important.

Refine the Model





The model's performance is evaluated using two key metrics: ROC-AUC and Precision-Recall AUC.
ROC-AUC Score: 0.82
The ROC-AUC score of 0.82 indicates that the model has a good ability to distinguish between the positive and negative classes. The closer this value is to 1, the better the model's performance. A score of 0.82 suggests that the model performs well in terms of overall classification accuracy across different thresholds.
Precision-Recall AUC: 0.45
The Precision-Recall AUC score of 0.45 suggests that the model's performance is moderate when specifically evaluating the trade-off between precision and recall. Since this score is closer to 0.5, it indicates that the model struggles to maintain both high precision and recall simultaneously, especially in a dataset that may be imbalanced.

Hyperparameter Tuning for Logistic Regression:

The model demonstrates strong overall discriminative power with a ROC-AUC of 0.82, but faces challenges in balancing precision and recall, as evidenced by the Precision-Recall AUC of 0.45. This discrepancy highlights the importance of considering both metrics, especially in imbalanced datasets.

Summary and Comparison with Baseline Model:
Despite tuning, the logistic regression model's performance metrics are virtually identical to the baseline model. The accuracy remains high due to the majority class (False), but the model fails entirely to identify the minority class (True), as indicated by the precision, recall, and F1-score of 0.00 for this class.
Key Points:
No Improvement: The tuning did not improve the model's ability to detect the minority class (True). The performance metrics for the minority class are the same as those in the baseline model.
Persistent Bias: The model remains biased towards the majority class, reflecting the ongoing issue of class imbalance.
Further Action Required: This indicates that the tuning strategy was insufficient to address the class imbalance. Further steps, such as more aggressive resampling, adjusting the class weights, or exploring different algorithms, are necessary to achieve better performance on the minority class.

Build and Evaluate a Decision Tree Model

Decision Tree Model Performance

Accuracy: 0.94

	precision	recall	f1-score	support
False	0.95	0.98	0.97	857
True	0.87	0.72	0.79	143
accuracy			0.94	1000
macro avg	0.91	0.85	0.88	1000
weighted avg	0.94	0.94	0.94	1000

The Decision Tree model outperforms the logistic regression models, reaching an accuracy of 0.94. It considerably improves the precision (0.87), recall (0.72), and F1-score (0.79) for the minority class. This shows that the Decision Tree is significantly more effective at recognising and correctly classifying occurrences of the minority class, giving it a more balanced and accurate model than the previously evaluated logistic regression models. The overall performance indicators, particularly for the minority class, are more favourable, making the Decision Tree a preferable choice for this classification assignment given the dataset's imbalance.

Tuned Decision Tree Model Performance

Accuracy: 0.94

	precision	recall	f1-score	support
False	0.95	0.98	0.97	857
True	0.85	0.72	0.78	143
accuracy			0.94	1000
macro avg	0.90	0.85	0.87	1000
weighted avg	0.94	0.94	0.94	1000

The tuned Decision Tree model maintains a high accuracy of 0.94, matching the performance of the previous model. The precision (0.85) and F1-score (0.78) for the True class are slightly lower than the earlier model (which had 0.87 precision and 0.79 F1-score). The recall for the True class remains consistent at 0.72. The differences are minor, indicating that while the tuning may have made slight adjustments, the overall performance remains robust and consistent with the earlier decision tree model. This suggests that the model is well-tuned, with only marginal trade-offs in precision and F1-score for the minority class.

Recommendations

Application of Predictions:

The decision tree model's strong performance suggests it is well-suited for use in identifying customers at risk of churning. This model can be deployed in scenarios where quick, real-time predictions are needed, such as during customer service interactions or targeted marketing campaigns aimed at customer retention.

However, the model may still struggle in edge cases where customer behavior is unusual or does not align with the patterns seen in the training data. Therefore, predictions should be interpreted with caution in such contexts, and the business should have contingency plans for dealing with misclassifications.

Business Strategy Adjustments:

Proactive Customer Retention: Based on the model’s predictions, the business could implement proactive retention strategies. For instance, customers identified as high-risk for churn could be offered personalized incentives, such as discounts or service upgrades.

Feature Modification: The business can experiment with modifying specific features to influence churn outcomes. For example, if the model identifies that customers with high customer service call rates are more likely to churn, the business could invest in improving customer service efficiency or providing alternative support channels to reduce the need for calls.

Targeted Interventions:

High Customer Service Call Rates:

- Insight: The Decision Tree model likely identified a high number of customer service calls as a significant predictor of churn. Customers who frequently contact customer service may be experiencing unresolved issues or dissatisfaction.
- Intervention: Implement a Proactive Customer Service Outreach Program. Customers with a high call frequency should be flagged for follow-up by a dedicated support team focused on resolving their issues before they escalate. This team can offer personalized solutions or compensation if necessary to improve customer satisfaction and reduce churn.

International Plan Subscribers with Low Usage:

- Insight: The model may have found that customers who subscribe to an international plan but do not frequently use international services are at higher risk of churning. These customers might feel that they are not getting value from their subscription.
- Intervention: Introduce a Personalized Usage Review for these customers, offering them tailored advice on how they can maximize the value of their plan. Alternatively, consider offering a downgraded or alternative plan with features that align more closely with their usage patterns, potentially at a lower cost. This approach can increase customer satisfaction and reduce the likelihood of them canceling their subscription.

Customers with High Daytime Minutes:

- Insight: Customers with high usage during the day might be key users of your service, but they may also be more price-sensitive or considering switching to competitors offering better rates or packages.
- Intervention: Develop a Loyalty Rewards Program that targets these high-usage customers. Offer them special discounts or bonus minutes if they commit to a long-term plan. Highlight the value and reliability of your service, potentially bundling additional features that are beneficial for high-usage customers, such as rollover minutes or unlimited daytime calling.

Low Usage of Voice Mail Plan:

- Insight: Customers who subscribe to a voice mail plan but seldom use it might feel that this feature is unnecessary, leading to dissatisfaction with their current package.
- Intervention: Provide a Plan Optimization Service that reviews customer usage patterns and suggests the most cost-effective plan. For those with low voice mail usage, offer alternative plans that exclude voice mail but include other features they might value more. This can help customers feel that they are getting better value for their money, increasing their loyalty.

Recent Changes in Usage Patterns:

- Insight: A sudden change in a customer's usage pattern, such as a significant drop in usage, might indicate that they are considering switching to another provider or that their needs are not being met.
- Intervention: Set up a Usage Pattern Monitoring System that triggers alerts when significant changes in usage are detected. This could be followed by a personalized communication offering tailored solutions, such as a check-in call or an exclusive offer to address their changing needs and prevent churn.

Customers with No Recent Plan Changes:

- Insight: Customers who have not made any changes to their plans for an extended period might be at risk of churn due to their plans no longer aligning with their current needs.
- Intervention: Implement a Customer Plan Review Campaign, where you proactively reach out to these customers, offering a free consultation to review their current plan. Provide recommendations or upgrades that better suit their current usage, ensuring that they feel valued and well-served.

FAUGET

THE END