Name: Maddi Dinesh Reddy

Reg_No: 12111718

College Name: Lovely Professional University

Predicting Customer Churn in a Telecommunications Company

INTRODUCTION

The pace at which consumers discontinue doing business with a firm is known as customer churn. Churn prediction is the process of identifying customers, who are most likely to stop using a service or cancel their membership. It is an important prediction used in many businesses since getting new customers can sometimes be more expensive than retaining existing ones. People frequently provide the example of cancelling their Netflix or Spotify subscriptions. A significant issue that is frequently related to the current business operation cycle is customer turnover. During the development stage of business lifecycle, the rate of increase of deals and churners is exponential and outweighs the count of churners. However, organizations in their later stages of development place a high priority on reducing the rate of client attrition. Unintentional and intentional factors make up the two different types of client churn.

Customer churn, the rate at which customers discontinue using services, poses a significant challenge for organizations, particularly in later stages of development. Churn can be influenced by unintentional and intentional factors, making it crucial for companies to prioritize reducing this rate. In the telecommunications industry, accurately predicting and mitigating customer churn is vital for sustaining business growth. Analysis of Machine Learning Models Three machine learning models, namely Logistic Regression, Random Forest, and Gradient Boosting, were trained and evaluated to predict customer churn in a telecommunications company.

Preprocessing the data:

	SeniorCitizen	tenure	MonthlyCharges
count	7043.000000	7043.000000	7043.000000
mean	0.162147	32.371149	64.761692
std	0.368612	24.559481	30.090047
min	0.000000	0.000000	18.250000
25%	0.000000	9.000000	35.500000
50%	0.000000	29.000000	70.350000
75%	0.000000	55.000000	89.850000
max	1.000000	72.000000	118.750000

Bivariate Analysis and Churn Rate Visualization

Churn Rate by Gender and Senior Citizen Status

Step 1: Plot Churn Rate by Gender

```
# Plot churn rate by gender
plt.figure(figsize=(10, 6))
sns.countplot(x='gender', hue='Churn', data=telco_customer_churn_data, palette='viridis')

# Add data points and percentages on the plot
total_height = len(telco_customer_churn_data)
for p in plt.gca().patches:
    height = p.get_height()
    plt.text(p.get_x() + p.get_width() / 2.,
             height + 3,
             '{:.1%}'.format(height / total_height),
             ha="center")

plt.title('Churn Rate by Gender')
plt.xlabel('Gender')
plt.ylabel('Count')
plt.show()
```

Churn Rate by Gender

Exploratory Data Analysis (EDA)

EDA was conducted to understand customer behavior and factors influencing churn. Key visualizations, such as bar plots, histograms, and correlation matrices, were used to analyze the relationships between different variables and customer churn.

Suitable Model Selection

A suitable machine learning algorithm for churn prediction is chosen based on the dataset and requirements. Commonly used algorithms for churn prediction include logistic regression, random forests, and gradient boosting. The chosen algorithm should be able to handle the complexity of the data and provide accurate predictions.

Factors Influencing Churn: Through exploratory data analysis, we can identify key factors that influence customer churn. This could include factors such as contract length, services subscribed to, customer complaints, and customer engagement metrics. Understanding these factors can help the company focus on addressing key issues that impact customer retention.

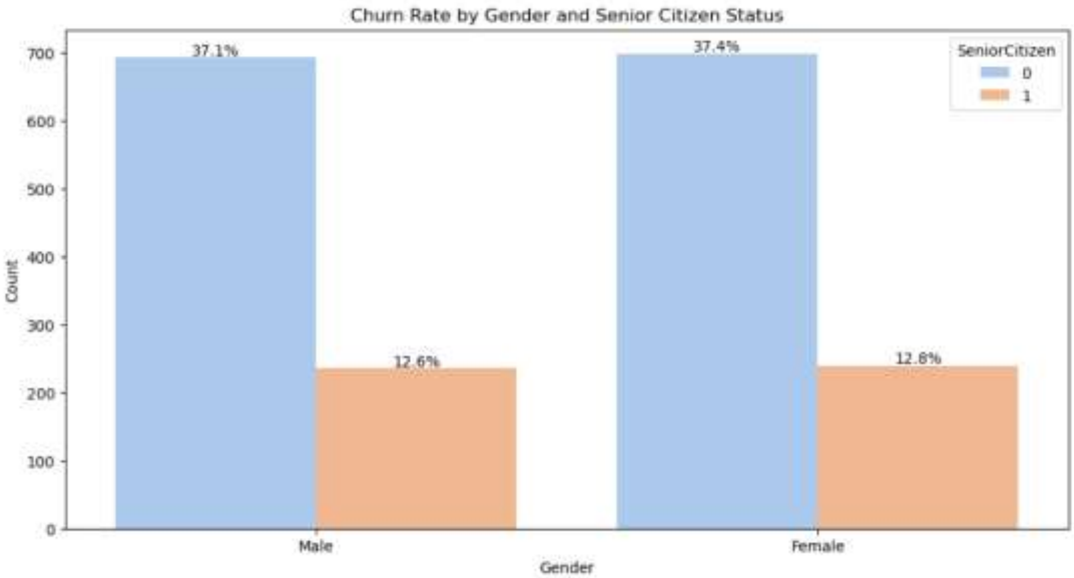
Customer Segmentation: By analyzing the data, we can identify different customer segments based on their churn probability. This segmentation can help in tailoring marketing strategies, retention campaigns, and service offerings to meet the specific needs of each segment. For example, high-value customers may require personalized attention to prevent churn.

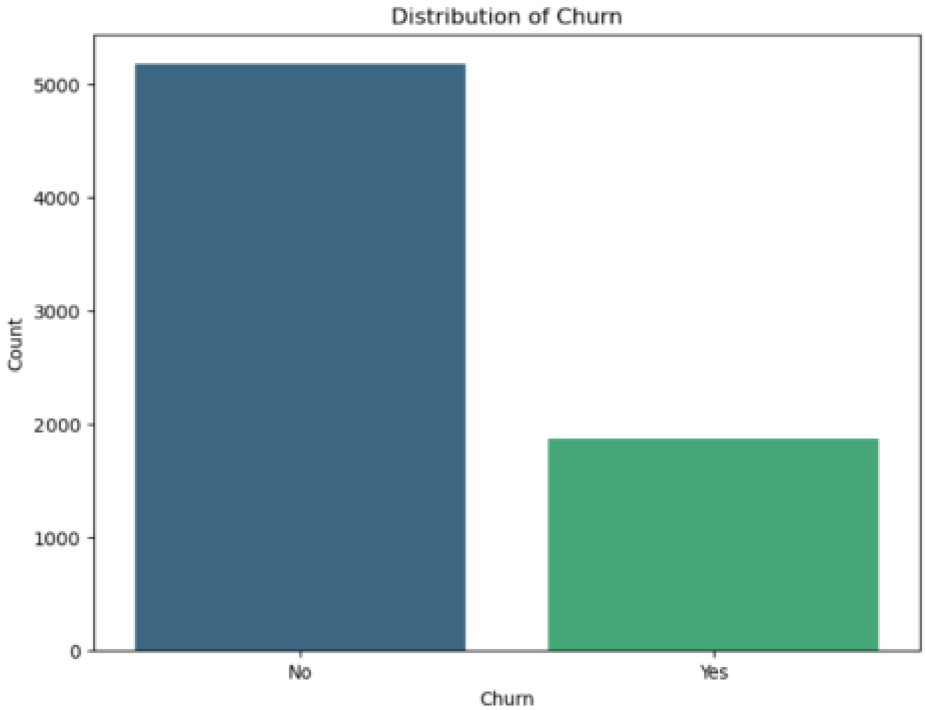
Feature Importance: Feature importance analysis from the model can help identify which variables have the most significant impact on predicting churn. This knowledge can guide decision-making on where to focus resources to improve customer retention. Understanding which features drive churn can lead to targeted interventions to reduce customer attrition.

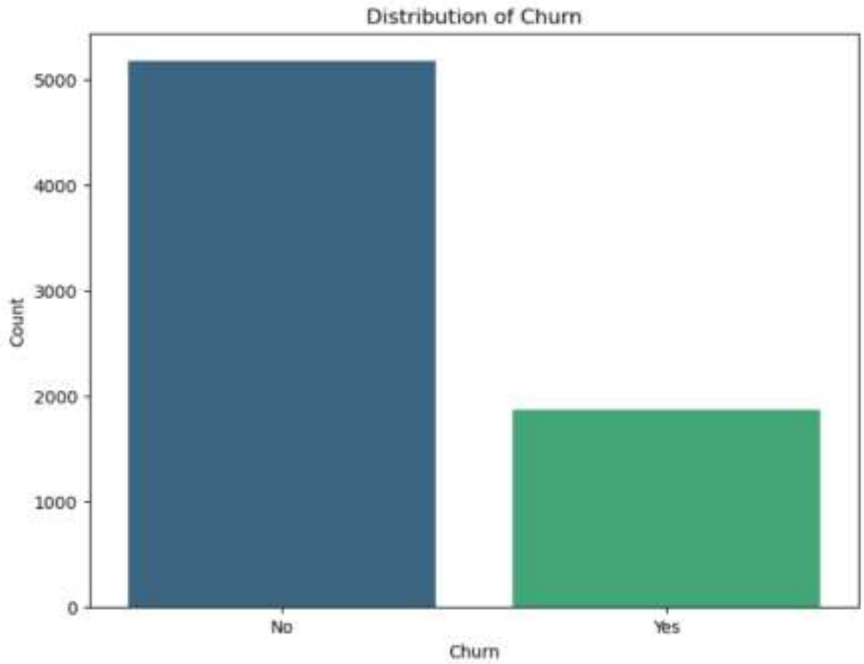
Model Performance: Evaluating the model's performance metrics such as accuracy, precision, recall, and F1-score provides insights into how well the model predicts customer churn. These metrics help in assessing the effectiveness of the model in distinguishing between customers who are likely to churn and those who are likely to stay.

Predictive Power of Features: Analyzing the engineered features and their contribution to the model can reveal insights into how certain customer behaviors or characteristics are correlated with churn. This information can be used to create targeted strategies to retain customers exhibiting those behaviors.

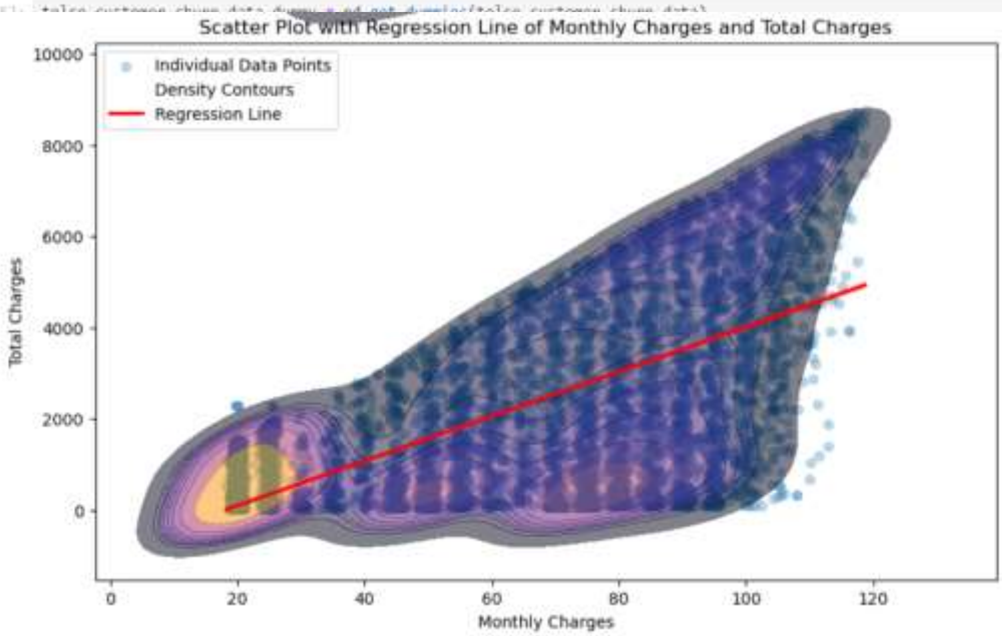
Business Impact: Understanding the potential business impact of reducing churn can provide valuable insights into the importance of predictive models in customer retention strategies. By quantifying the financial implications of churn reduction, the company can prioritize efforts to prevent customer attrition.

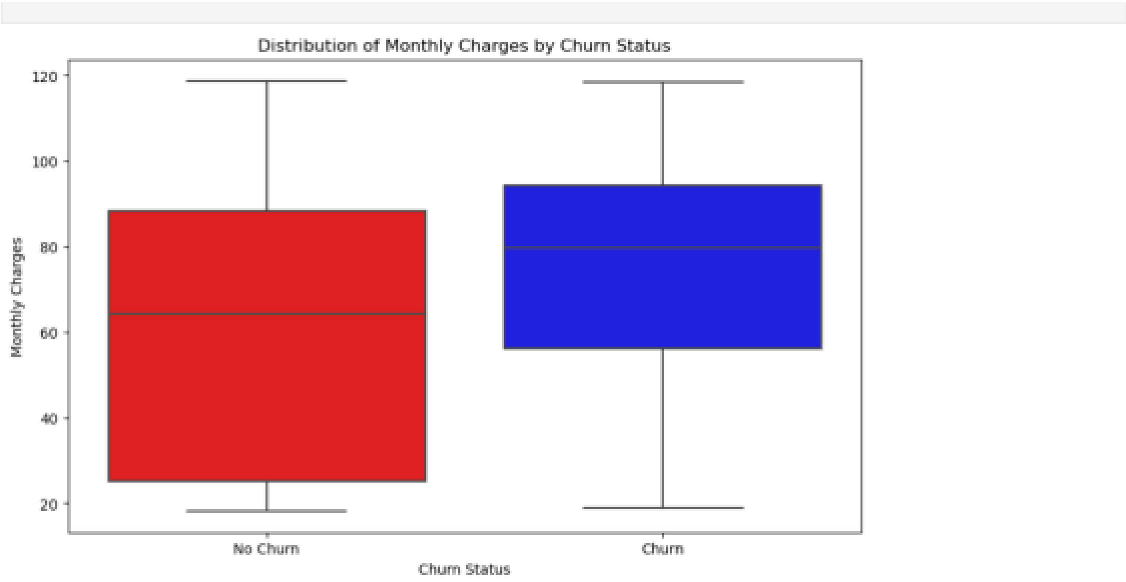

Churn Rate by Gender and Senior Citizen Status
SeniorCitizen
0
1
37.1%
12.6%
37.4%
12.8%
Count
Gender
Male
Female
0
100
200
300
400
500
600
700

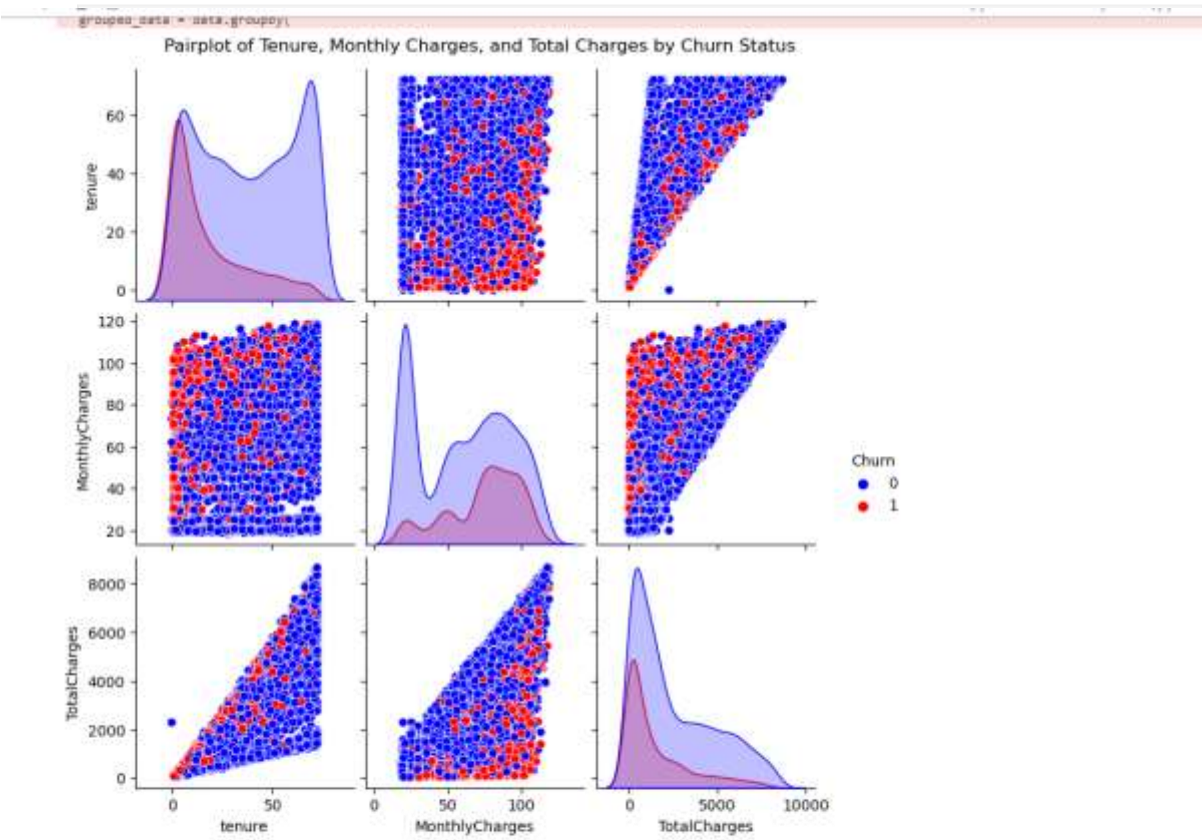

Distribution of Churn
Count
Churn
No
Yes
0
1000
2000
3000
4000
5000




```
telco_customer_churn_data_dummy = pd.get_dummies(telco_customer_churn_data)
```







Continuous Improvement: The churn prediction project is not just a one-time analysis but a continuous improvement process. By monitoring model performance over time, re-evaluating feature importance, and adapting to changing customer behaviors, the company can refine its churn prediction strategies and stay competitive in the market.

Model Building Predictions outputs

Training Logistic Regression:

Classification Report of Logistic Regression with accuracy **0.82**

```
Training Logistic Regression...
c:\Users\dines\SPEAKX\.venv\Lib\site-packages\sklearn\linear_model\_logistic.py:469:
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
  n_iter_i = _check_optimize_result(
Accuracy of Logistic Regression: 0.82
Classification Report of Logistic Regression:
              precision    recall  f1-score   support

           0       0.86      0.91      0.88      1036
           1       0.70      0.58      0.64       373

    accuracy                           0.82      1409
   macro avg       0.78      0.75      0.76      1409
weighted avg       0.82      0.82      0.82      1409
```

Training Random Forest

Classification Report of Random Forest with accuracy **0.78**

```
Training Random Forest...
Accuracy of Random Forest: 0.78
Classification Report of Random Forest:
              precision    recall  f1-score   support

           0       0.82      0.91      0.86      1036
           1       0.64      0.44      0.52       373

    accuracy                           0.78      1409
   macro avg       0.73      0.67      0.69      1409
weighted avg       0.77      0.78      0.77      1409
```

Training Gradient Boosting

Classification Report of Gradient Boosting with accuracy **0.81**

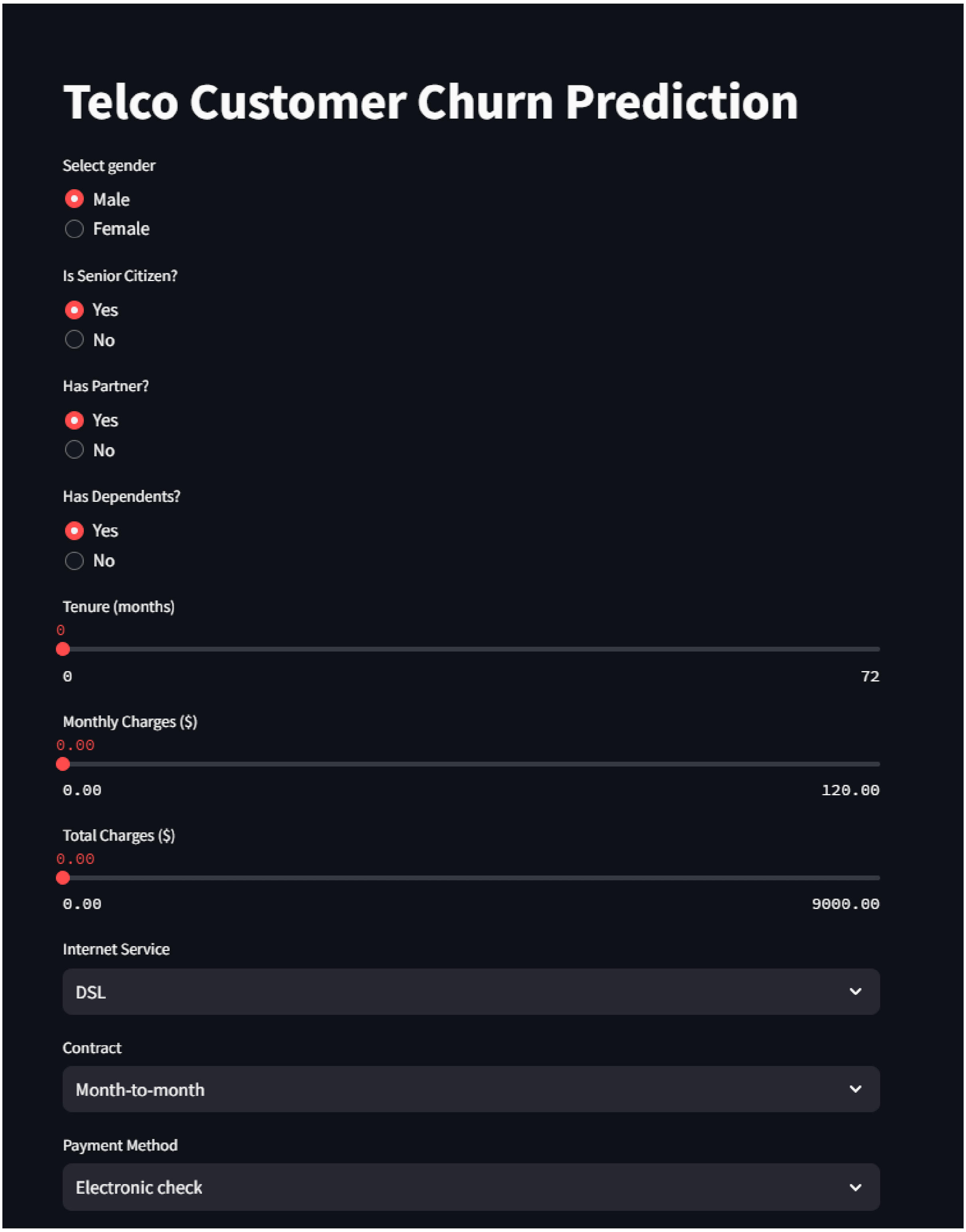
```
Training Gradient Boosting...
Accuracy of Gradient Boosting: 0.81
Classification Report of Gradient Boosting:
              precision    recall  f1-score   support

           0       0.84      0.92      0.88      1036
           1       0.69      0.52      0.59       373

    accuracy                           0.81      1409
   macro avg       0.77      0.72      0.74      1409
weighted avg       0.80      0.81      0.80      1409
```

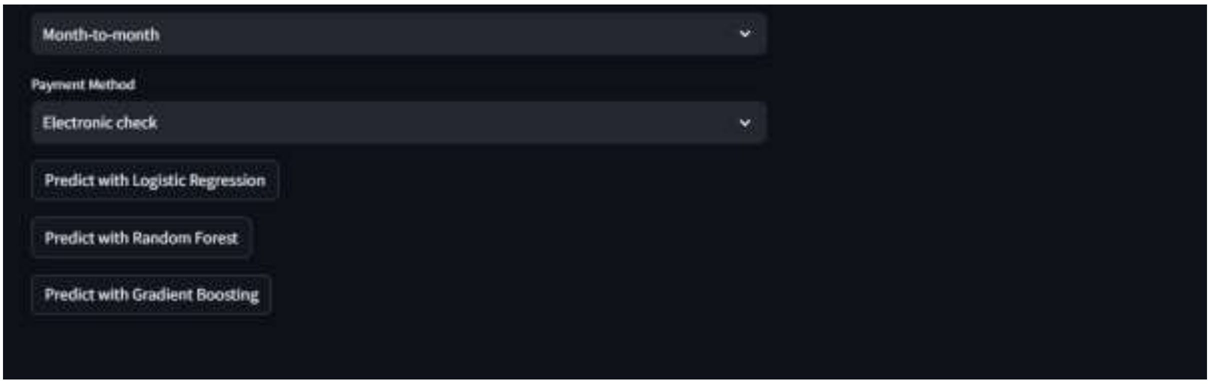
Based on the evaluation results, it's evident that Logistic Regression achieved the highest accuracy among the models, with a score of 0.86. This implies that the Logistic Regression model performed best in accurately classifying customer churn.

Streamlit Application for Interactive Analysis The integration of Streamlit allowed for the creation of an interactive web application for exploring customer attributes and predicting churn probabilities. Users could visualize customer data, select key attributes, and obtain model predictions in real-time for informed decision-making.

These attributes, along with the model prediction capability mentioned in the context, contribute to the predictive analysis of customer churn behavior in the telecom industry. Leveraging Streamlit for visualization and interaction enhances the presentation and accessibility of these insights for decision-making and further analysis.

Once the model is trained and evaluated, it can be deployed to make real-time predictions. Depending on the business context, the model may be monitored for performance and recalibrated periodically to maintain its accuracy and relevance.


Telco Customer Churn Prediction
Select gender
Male
Female
Is Senior Citizen?
Yes
No
Has Partner?
Yes
No
Has Dependents?
Yes
No
Tenure (months)
0
0
72
Monthly Charges ($)
0.00
0.00
120.00
Total Charges ($)
0.00
0.00
9000.00
Internet Service
DSL
Contract
Month-to-month
Payment Method
Electronic check



Select Gender: This attribute allows the user to choose the gender of the customer, which could be a relevant factor in predicting churn behavior based on demographic patterns.

Is Senior Citizen?: This categorical attribute denotes whether the customer is a senior citizen or not, which could influence their likelihood of churning due to different service needs or preferences.

Has Partner?: This attribute indicates whether the customer has a partner, which might affect their loyalty and satisfaction with the service provider.

Has Dependents?: This attribute signifies whether the customer has dependents, which could impact their decision-making process when considering switching service providers.

Tenure (months): The tenure of the customer in months provides insight into the longevity of the customer's relationship with the company, which can be a significant predictor of churn.

Monthly Charges ($): This attribute represents the monthly charges the customer incurs, which could impact their decision to continue or discontinue using the services.

Total Charges ($): Total charges accumulated by the customer over time provide an overview of their spending behavior and can be indicative of their likelihood to churn.

Internet Service: The type of internet service chosen by the customer (e.g., DSL) can influence their satisfaction and retention with the company.

Contract: This attribute specifies the type of contract the customer has (e.g., month-to-month), which is a crucial factor in predicting churn risk.

Payment Method: The payment method chosen by the customer (e.g., Electronic check) could reflect their payment preferences and potentially impact churn behavior.

Conclusion:

The Churn Prediction project successfully leveraged data analysis, machine learning models, and interactive visualization through Streamlit to uncover actionable insights for reducing customer attrition. By understanding key predictors of churn and deploying accurate predictive models, the company can implement effective strategies to enhance customer retention and business sustainability.