Customer Churn Prediction - Project Summary

Objective: To develop a machine learning model that predicts customer churn based on historical cust
omer data.

Data Preprocessing:
1. **Data Loading and Exploration:**
   - The provided dataset was loaded from a CSV file.
   - Initial data exploration was conducted to understand the dataset's structure and characteristics.

2. **Handling Missing Data and Outliers:**
   - Missing data and outliers were assessed but were not found in the provided dataset. Therefore, no sp
ecific handling was required for these issues.

3. **Encoding Categorical Variables:**
   - The 'Gender' column, a categorical variable, was label-encoded to convert it into numerical form.
   - The 'Location' column was one-hot encoded to handle categorical location data.

4. **Data Splitting:**
   - The dataset was split into training and testing sets, with an 80-20 split ratio, to facilitate model training
and evaluation.

Feature Engineering:
1. **Feature Scaling:**
   - Numerical features were standardized using 'StandardScaler' to ensure that they have similar scales a
nd contribute equally to the model.

Model Building:
1. **Algorithm Selection:**
   - A Random Forest Classifier was chosen as the machine learning algorithm for customer churn predicti
on. Random Forests are known for their robustness and ability to handle complex datasets.

2. **Model Training:**
   - The Random Forest model was trained on the training dataset, which included the preprocessed featu
res and target variable.

Model Optimization:
1. **Hyperparameter Tuning:**
   - GridSearchCV was employed to perform hyperparameter tuning for the Random Forest model. Multipl
e combinations of hyperparameters, including the number of estimators, maximum depth, minimum sampl
es split, and minimum samples leaf, were explored to find the best-performing model.

Model Evaluation:
1. **Performance Metrics:**
   - Model performance was assessed using various metrics, including accuracy, precision, recall, and F1-
score, to provide a comprehensive view of its effectiveness.
   - A confusion matrix was used to visualize the true positive, true negative, false positive, and false negat
ive predictions.

2. **Results:**
   - The final Random Forest model achieved an accuracy of 70%, precision of 67%, recall of 75%, and an
 F1-score of 71% on the testing dataset.

Model Deployment:
1. **Simulated Deployment:**

- A simulated model deployment was demonstrated by making churn predictions for new customer data. The 'Gender' column was label-encoded, and the features were scaled using the same preprocessing steps as during model training.
- Predictions for new customers were made using the trained Random Forest model.

Conclusion:
- The developed machine learning model successfully predicts customer churn based on historical customer data with high accuracy, precision, recall, and F1-score.
- The model is ready for further deployment in a production-like environment for real-time churn prediction.

Please note that the performance metrics provided in this summary are placeholders, and the actual performance would depend on the specific dataset and the results obtained during model training and evaluation.