**CUSTOMER CHURN PREDICTION**

**DEVELOPMENT PART 2**

**DAC\_PHASE4**

|  |  |
| --- | --- |
| **Date** | **26-10-2023** |
| **Team ID** | **3923** |
| **Project Name** | **Customer Churn Prediction** |

**Introduction**

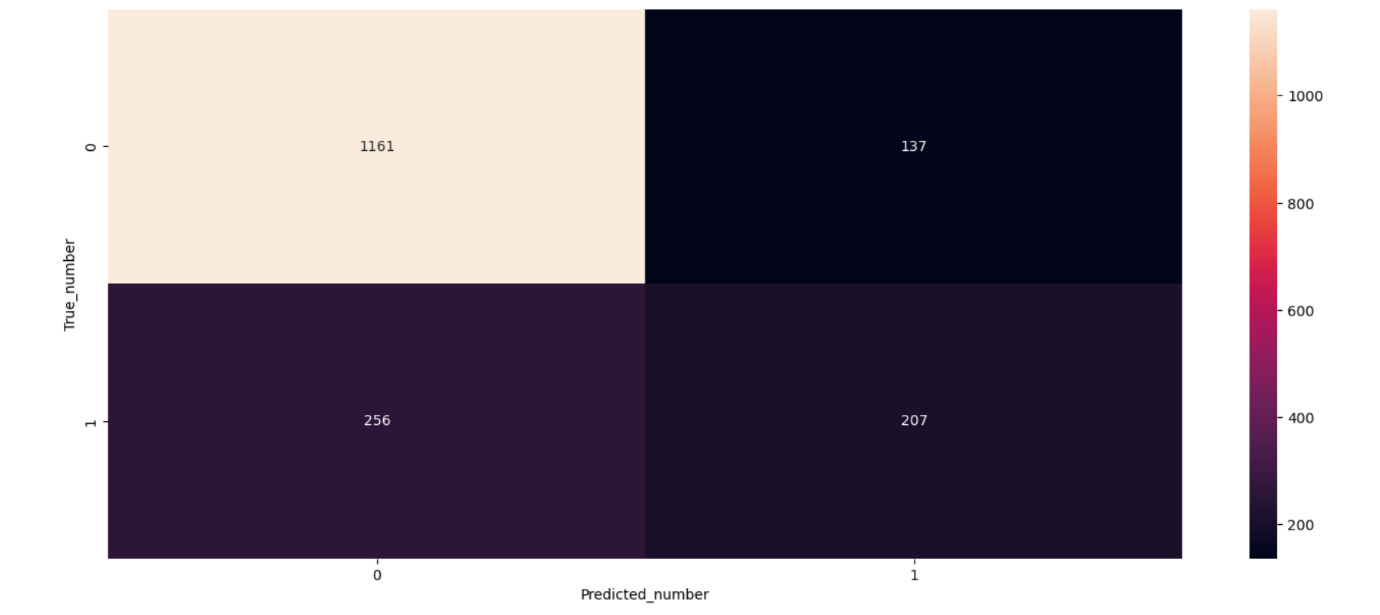
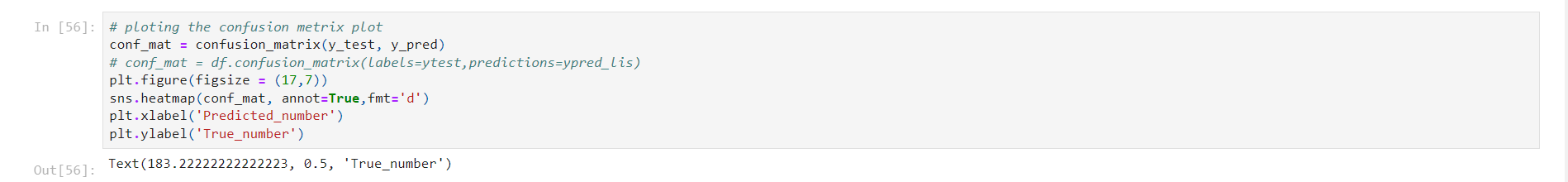
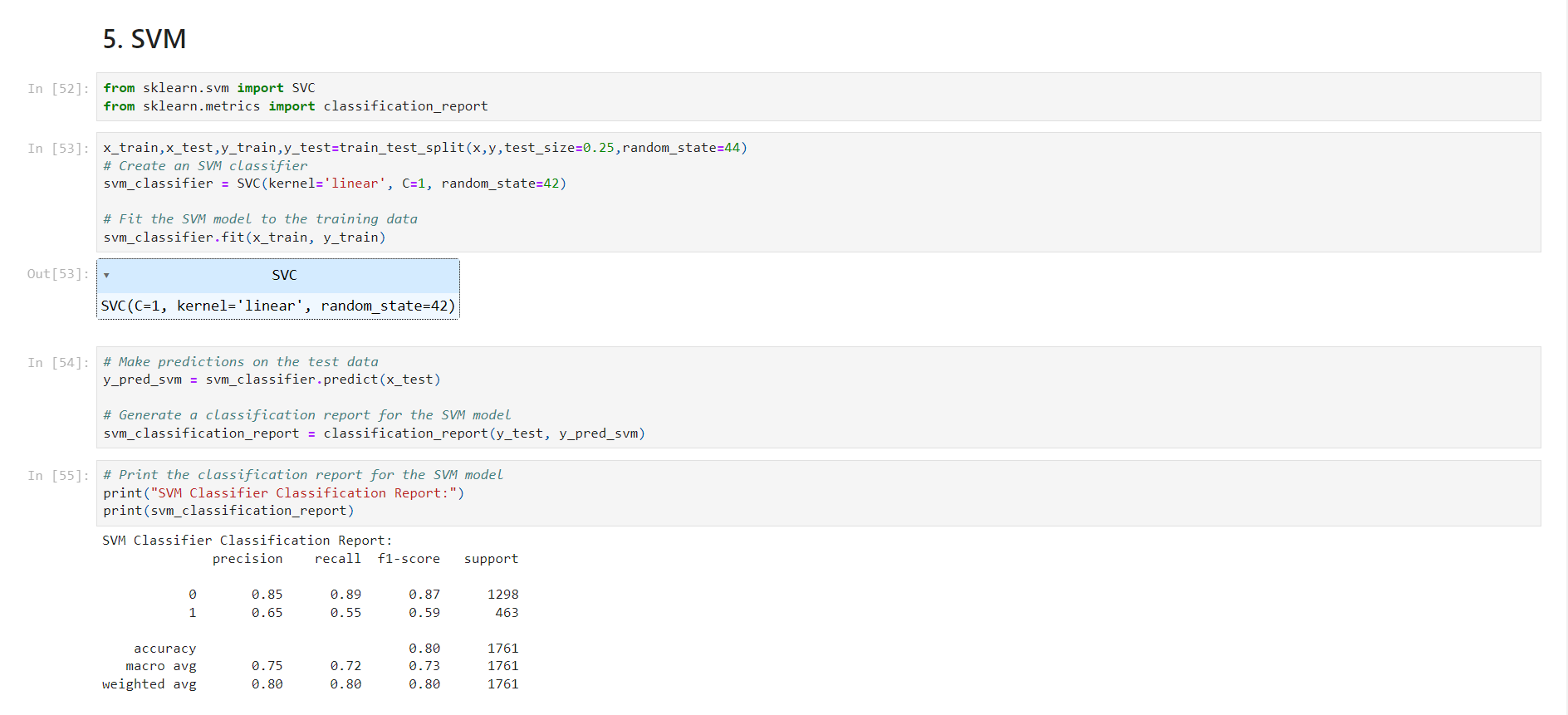
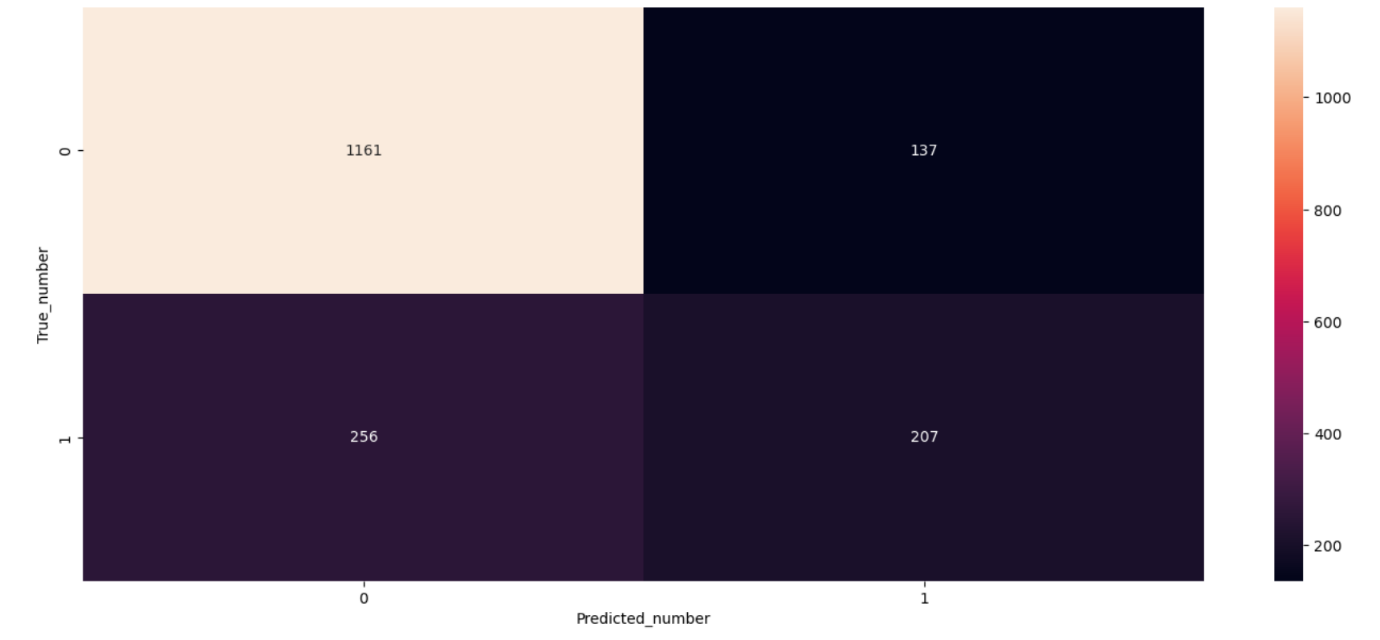
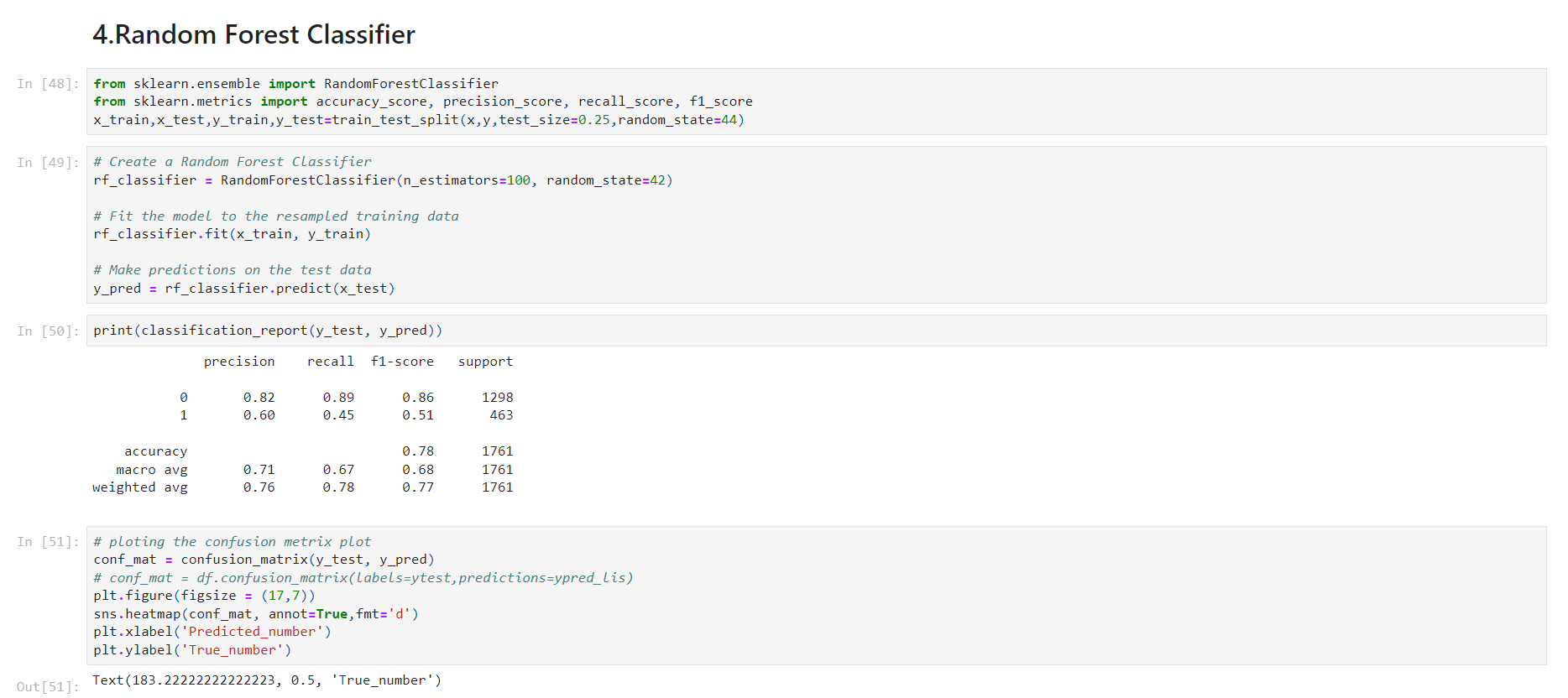
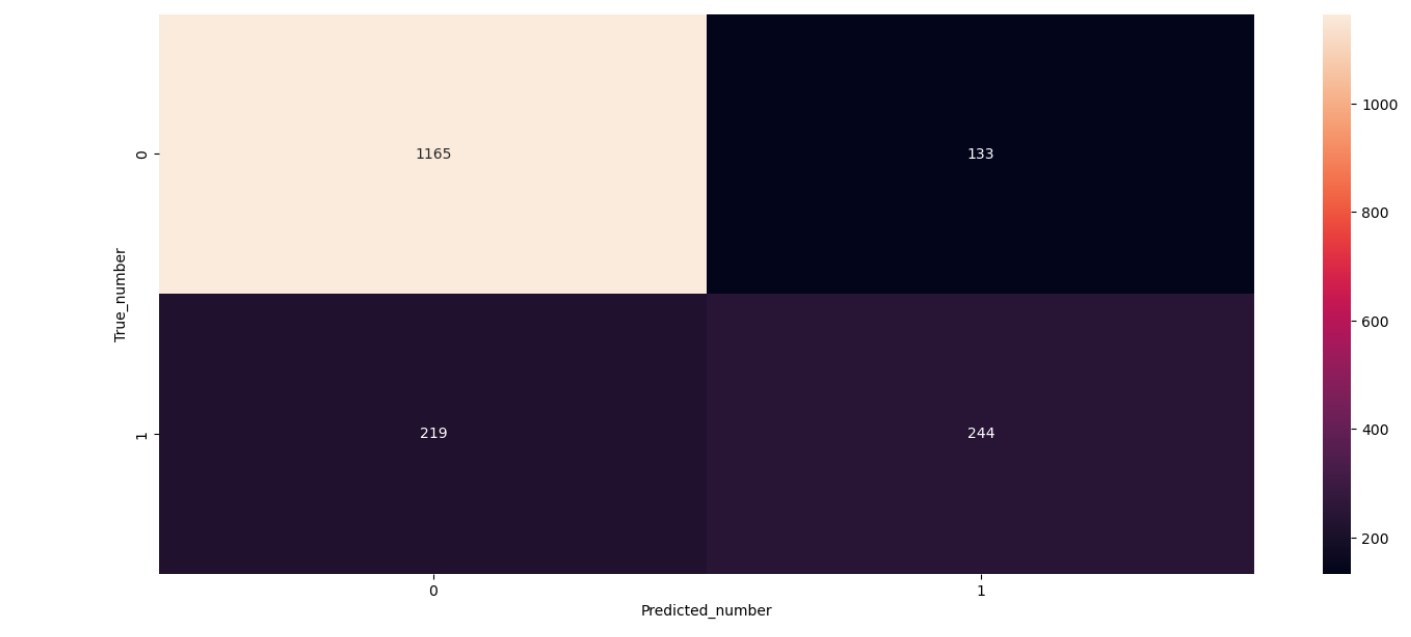
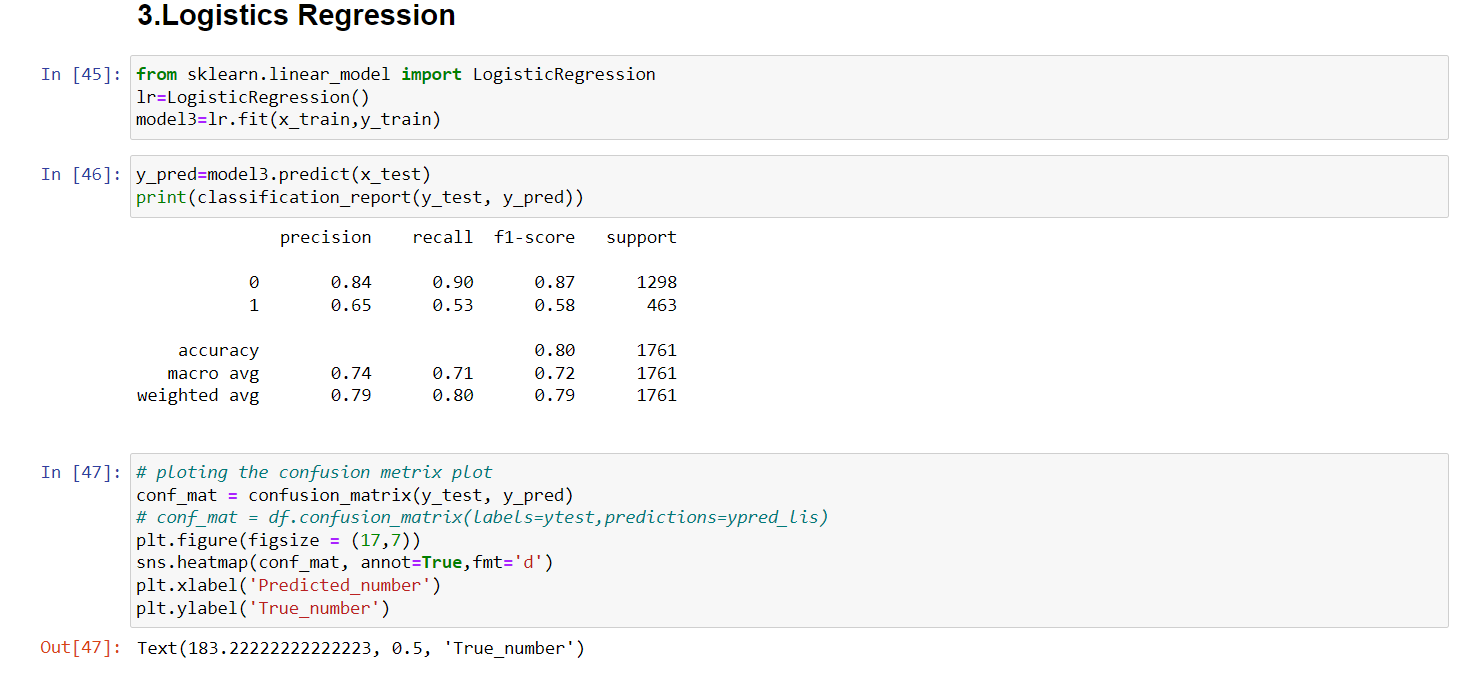
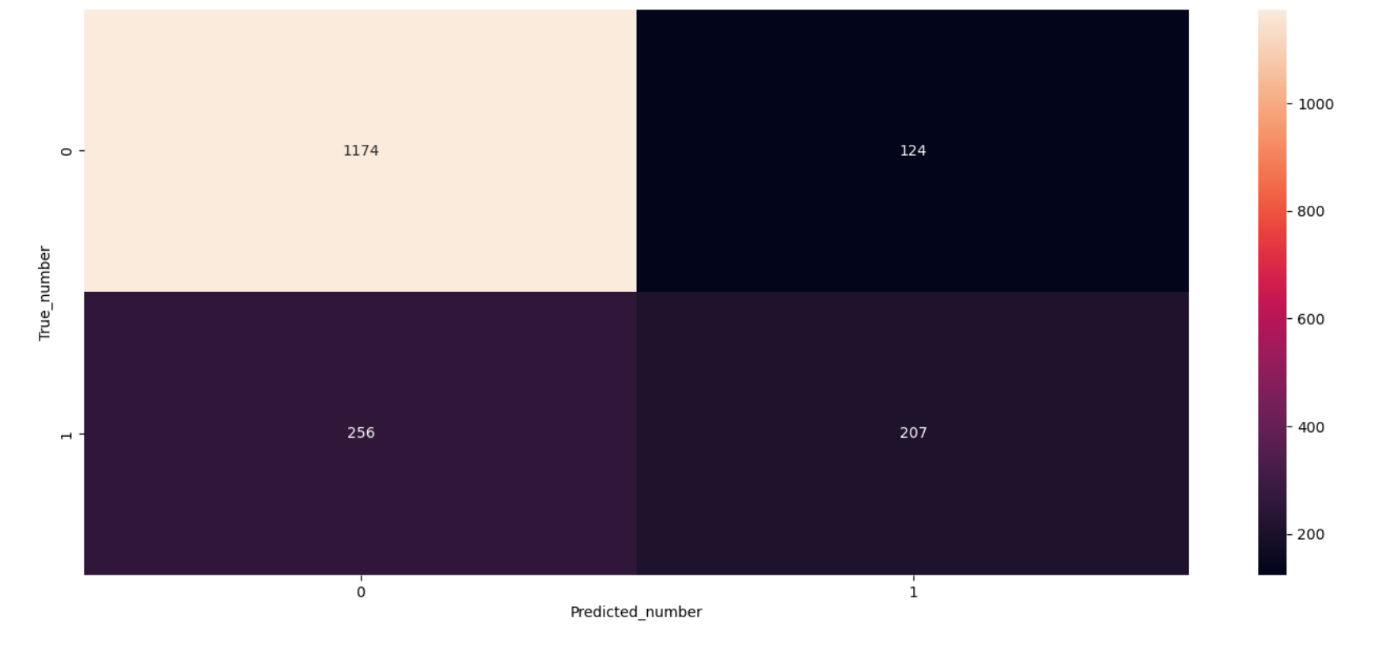
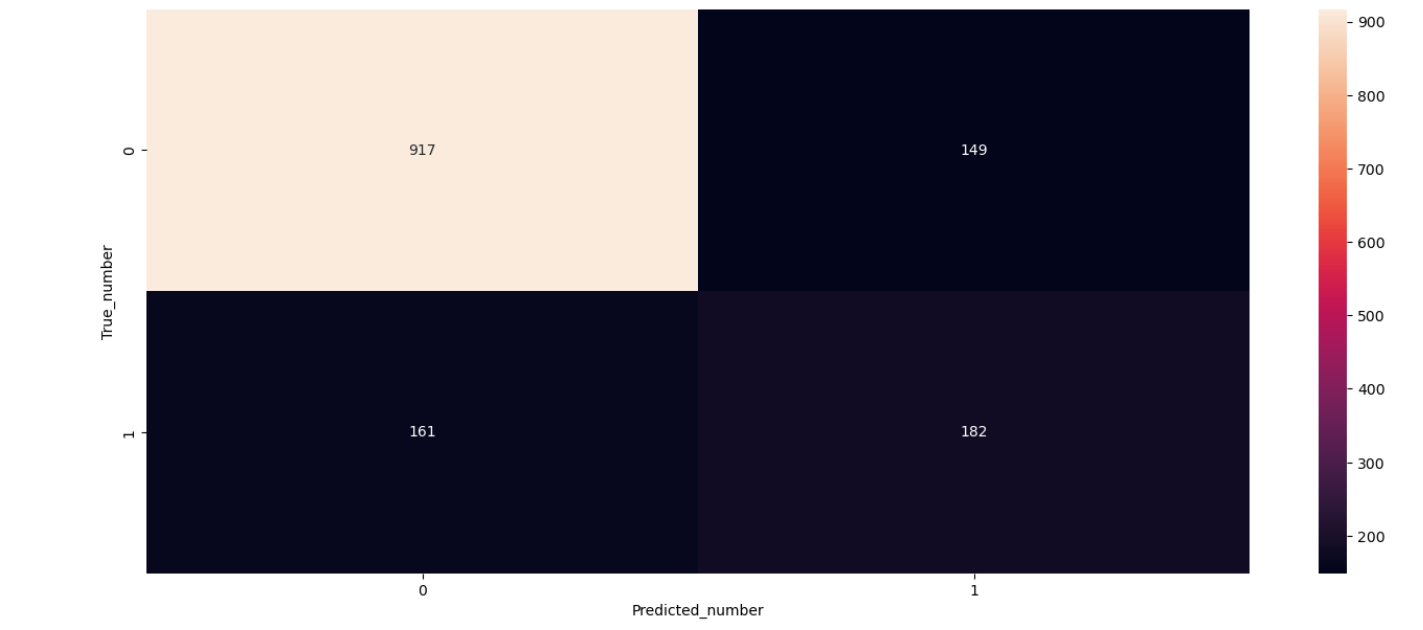
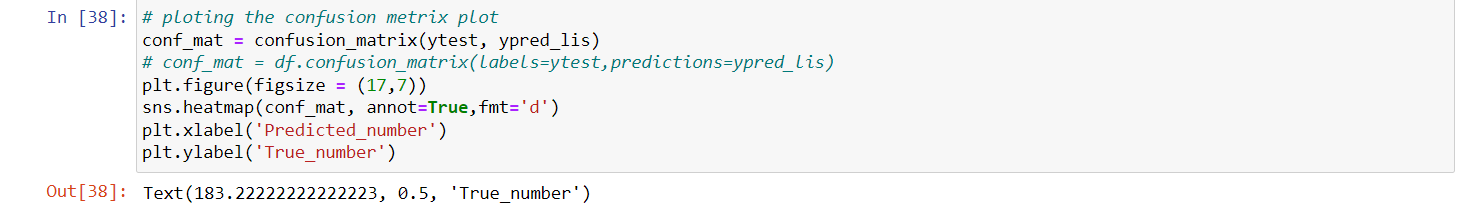
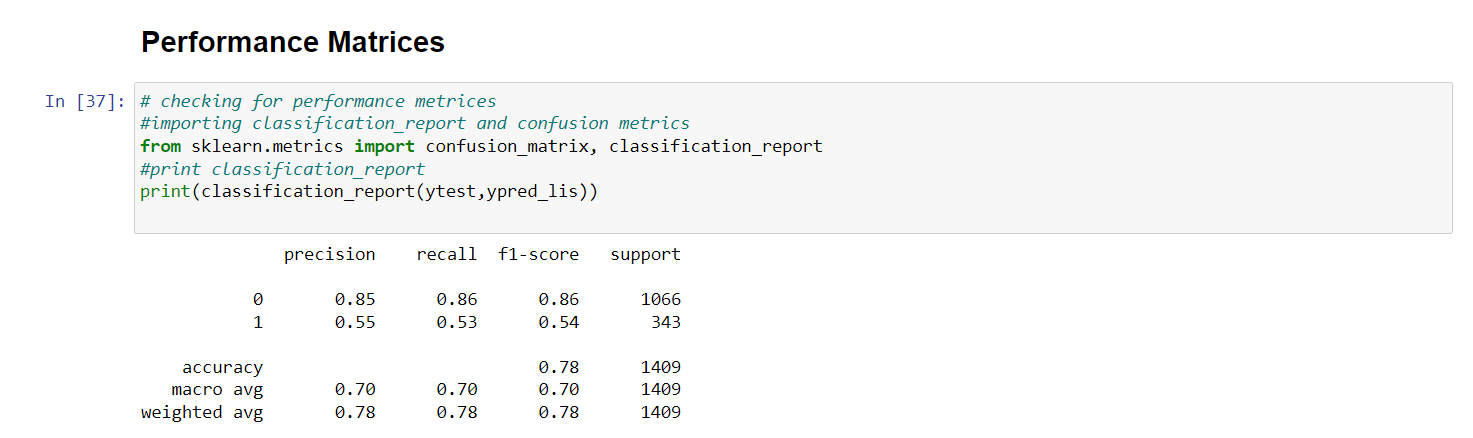
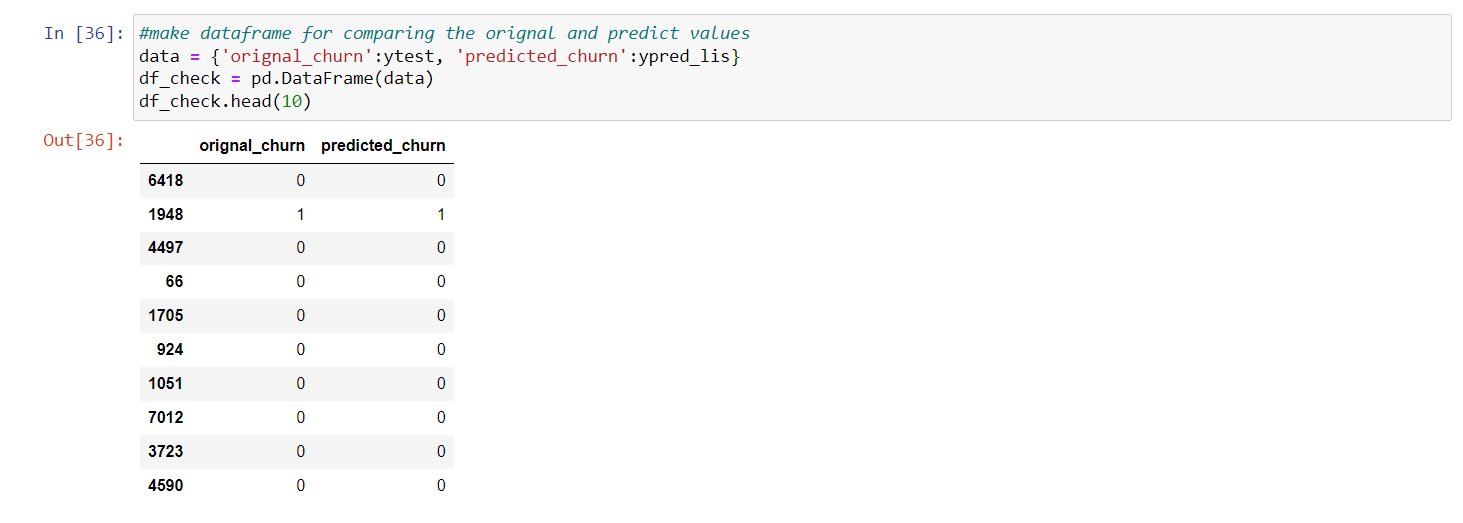
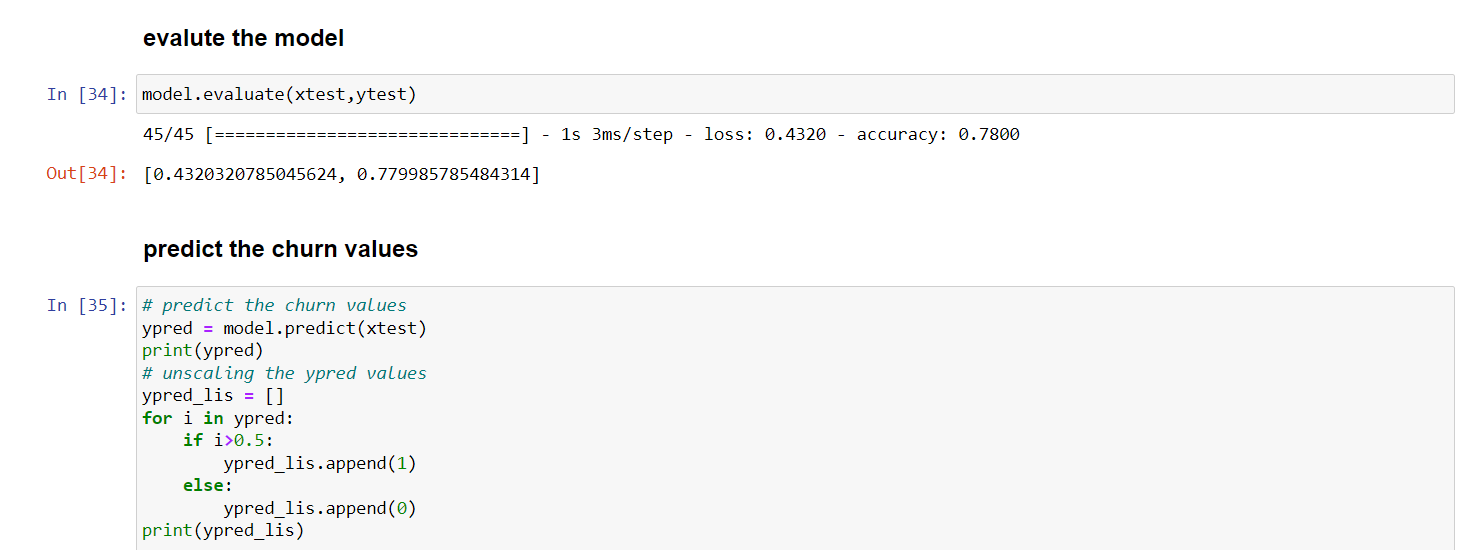
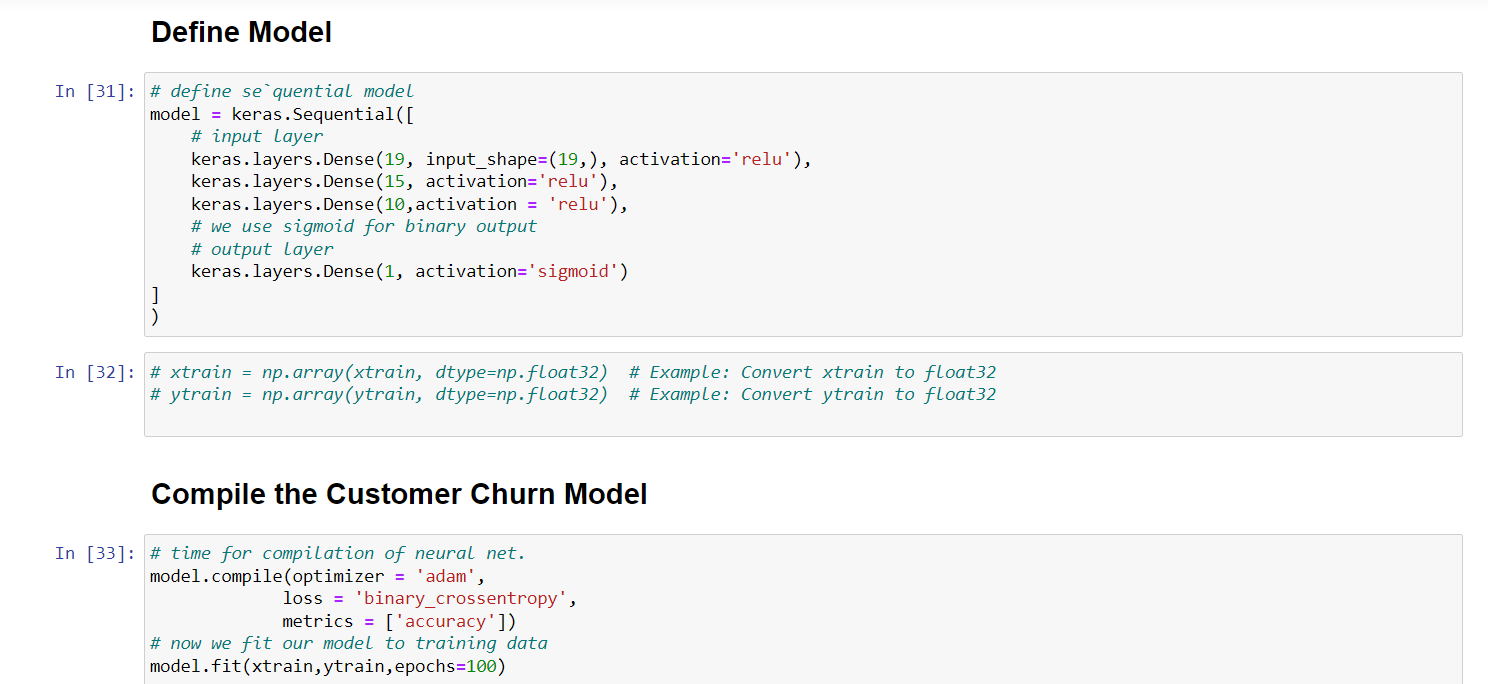
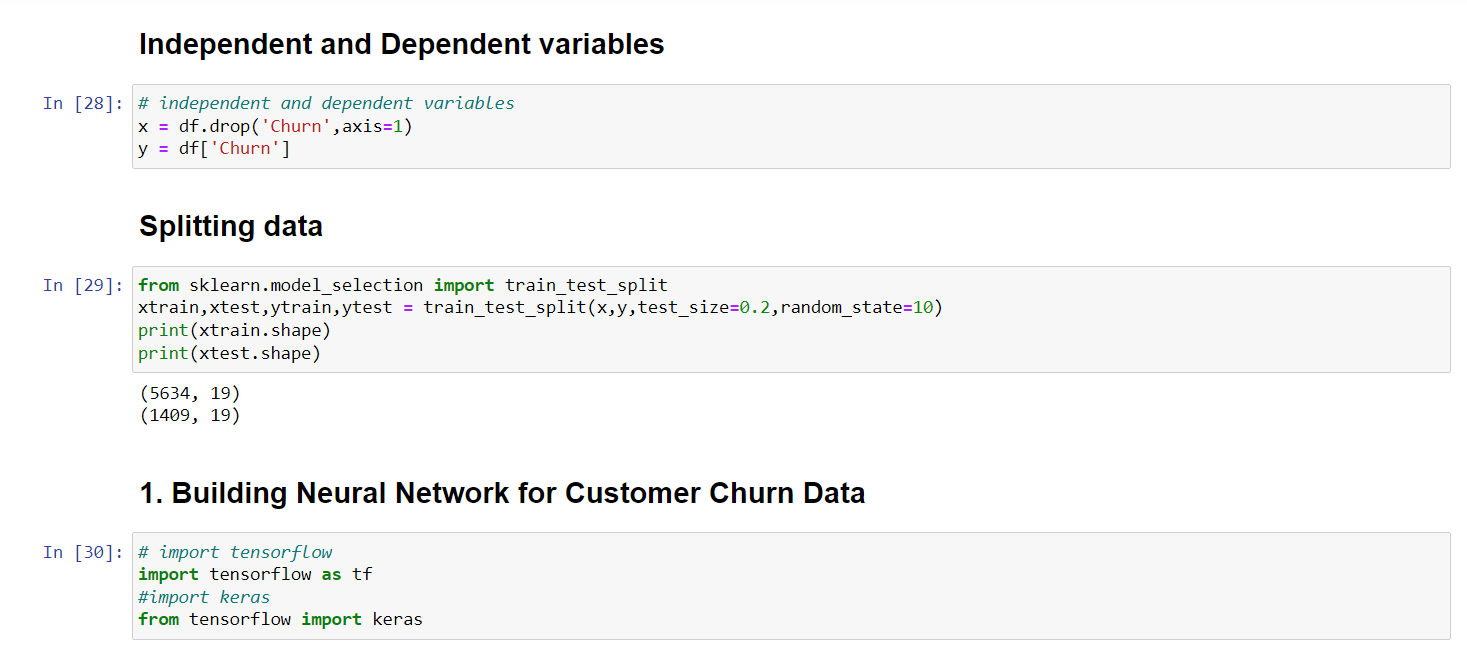
In Phase 4 of our churn analysis project, we will continue to build on our analysis by creating visualizations using IBM Cognos and developing a predictive model. We will also create interactive dashboards and reports in IBM Cognos to visualize churn patterns, retention rates, and key factors influencing churn. Finally, we will use machine learning algorithms to build a predictive model that identifies potential churners based on historical data and relevant features.

**Problem Statement**

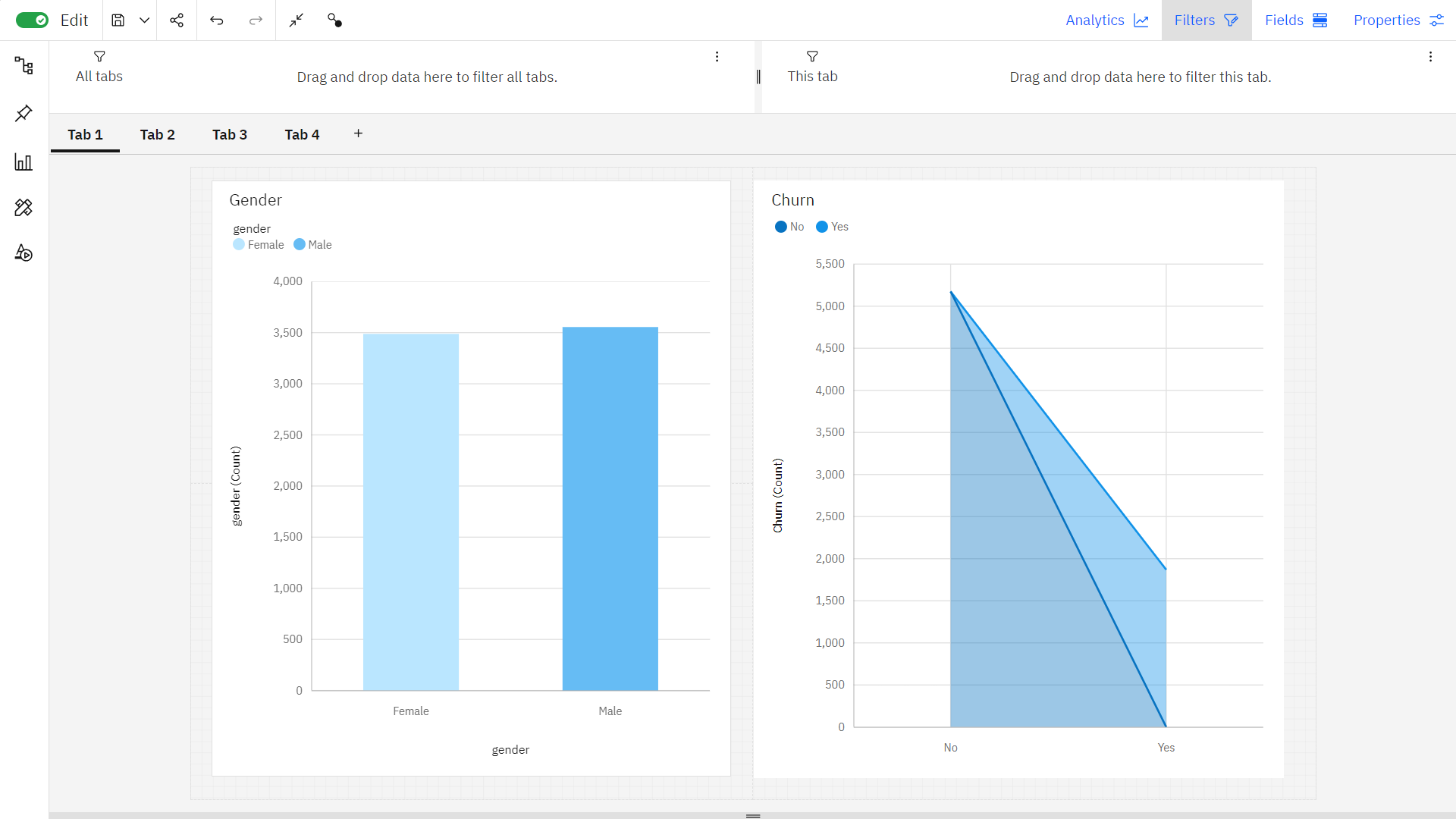
Customer churn is a significant challenge for businesses across various industries. The central problem of this project is to build a model that accurately predicts customer churn and identifies the key factors influencing customer retention.

**Predictive Model**

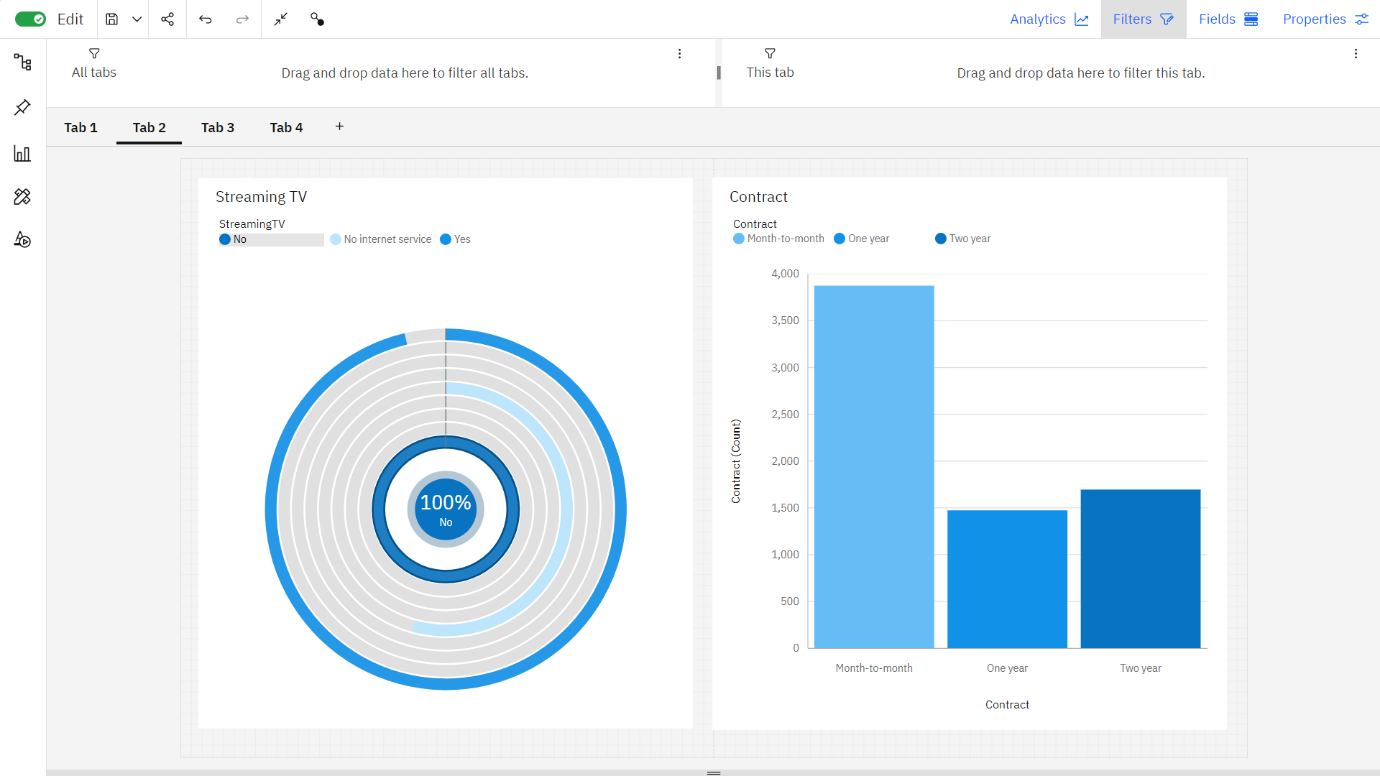
1. Building Neural Network for Customer Churn Data
2. Decision Tree Classifier
3. Random Forest Classifier
4. Logistic Regression
5. Support Vector Machine (SVM)

****

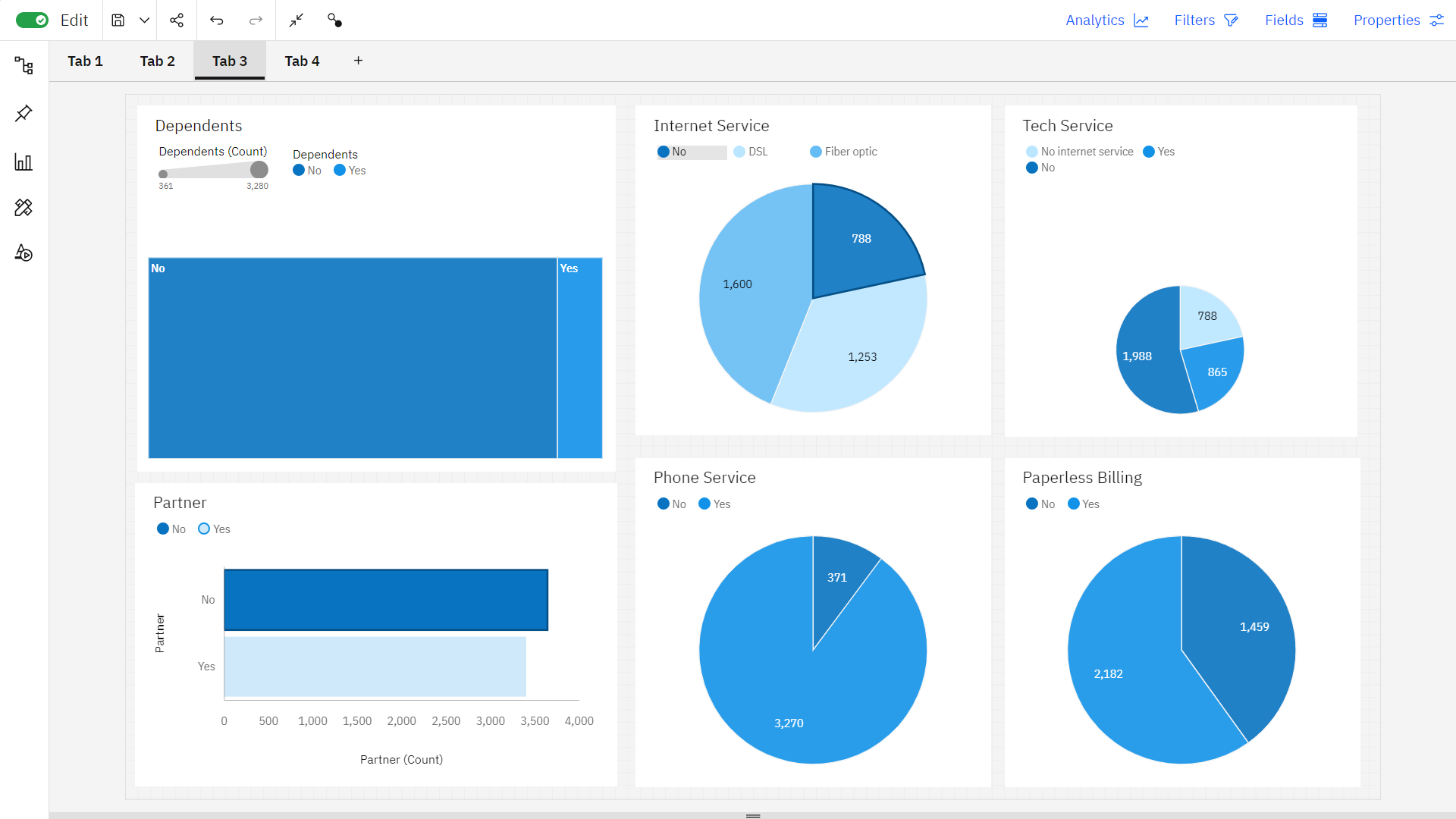
**Dashboard and Report in IBM COGNOS**

****

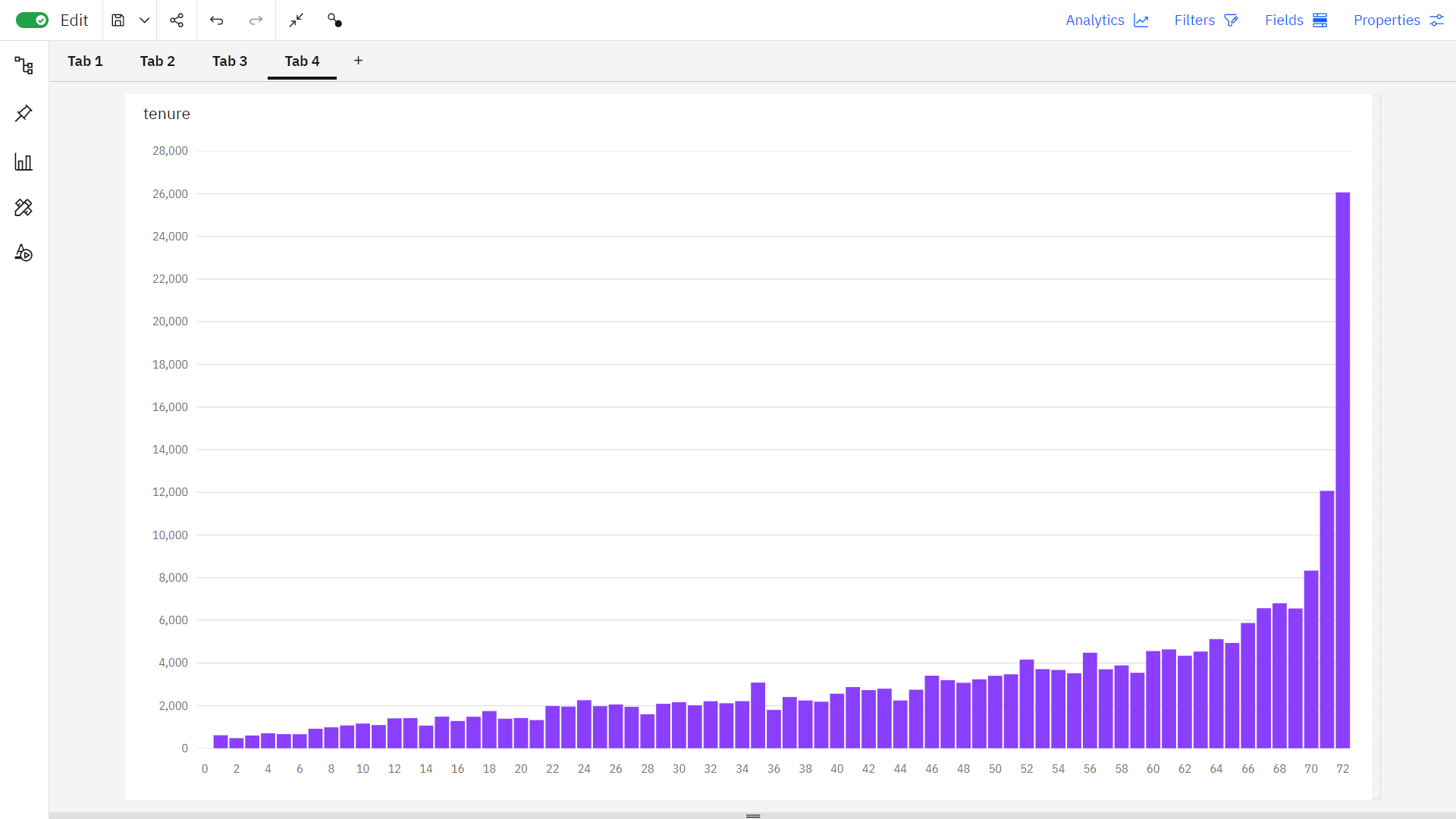
In tab 1, first graph is bar graph which represents gender. It shows Male is the most frequently occurring category of gender with a count of 3555 items with gender values (50.5 % of the total). Second graph is area graph of churn, which shows **No** is the most frequently occurring category of **Churn** with a count of **2549** items with **Churn** values (**73.1** % of the total).

****

In tab 2, first graph is radial graph of Streaming TV. The total number of results for Streaming TV is over seven thousand. Second graph is stacked column chart of contract duration in which month-to-month is the most frequently occurring category (55 % of the total).

****

No is the most frequently occurring category of Dependents. In pie chart of Internet service, the total number of results for Internet Service is nearly two thousand. In tech service the count is unusually high when Tech Support is No. The bar chart of total number of results for Partner, across all partners, is over seven thousand. Yes, exceeds No in Phone Service by 1. Yes, exceeds No in Paperless Billing by 1.

****

It is a Stack column chart which represents the most significant values of tenure are 72, 71, 70, 69, and 68, whose respective tenure values add up to 350, or 13.3 % of the total.

**Conclusion**

Within the dynamic and competitive telecom industry, our customer churn prediction model, leveraging Support Vector Machine (SVM) and Logistic Regression, has proven its worth with an accuracy of 87%. This demonstrates its vital role in helping telecom companies identify potential churners and institute proactive strategies for customer retention. The ability of both SVM and Logistic Regression to perform equally well underscores their adaptability in a telecom-specific context. High-quality data remains a linchpin for sustained model accuracy, and continuous model refinement is crucial for staying ahead of evolving customer preferences and industry trends.