Bank Customer Churn Analysis and Prediction

Madhava Krishnan N S

Mar 10, 2024

www.linkedin.com/in/madhava-krishnan-n-s

Outline

- Introduction
- Methodology
- Analysis Results
- Prediction Results
- Conclusion

Introduction

- This project delves into the intricate patterns and predictive aspects that drive customer attrition. Leveraging Power BI for insightful visualization and Jupyter notebooks for robust modeling, our methodology navigates through data wrangling, feature engineering, and meticulous model selection.

- Join us on a journey to uncover the dynamics influencing customer behavior and loyalty.

Methodology – Analysis Methodology

Data Collection:

- Downloaded customer churn data from Kaggle. https://www.kaggle.com/datasets/barelydedicated/bank-customer-churn-modeling
- Imported and loaded the data into Power BI for analysis.

Data Wrangling:

- Conducted data cleaning by replacing the values with appropriate names.
- Grouped continuous data, such as age, credit score, and bank balance, into relevant categories for a more comprehensive analysis.

Visualization:

- Created a comprehensive dashboard in Power BI that incorporates all features pertinent to customer churn.

The Power BI dashboard is accessible through the following link, providing a visual representation of the analyzed data.
https://github.com/Madhavananalyst/datascience/blob/main/Customer%20Churn%20Analysis.pbix

Methodology – Prediction Methodology

Data Collection:
- Loaded the relevant data into a Jupyter notebook for predictive analysis.

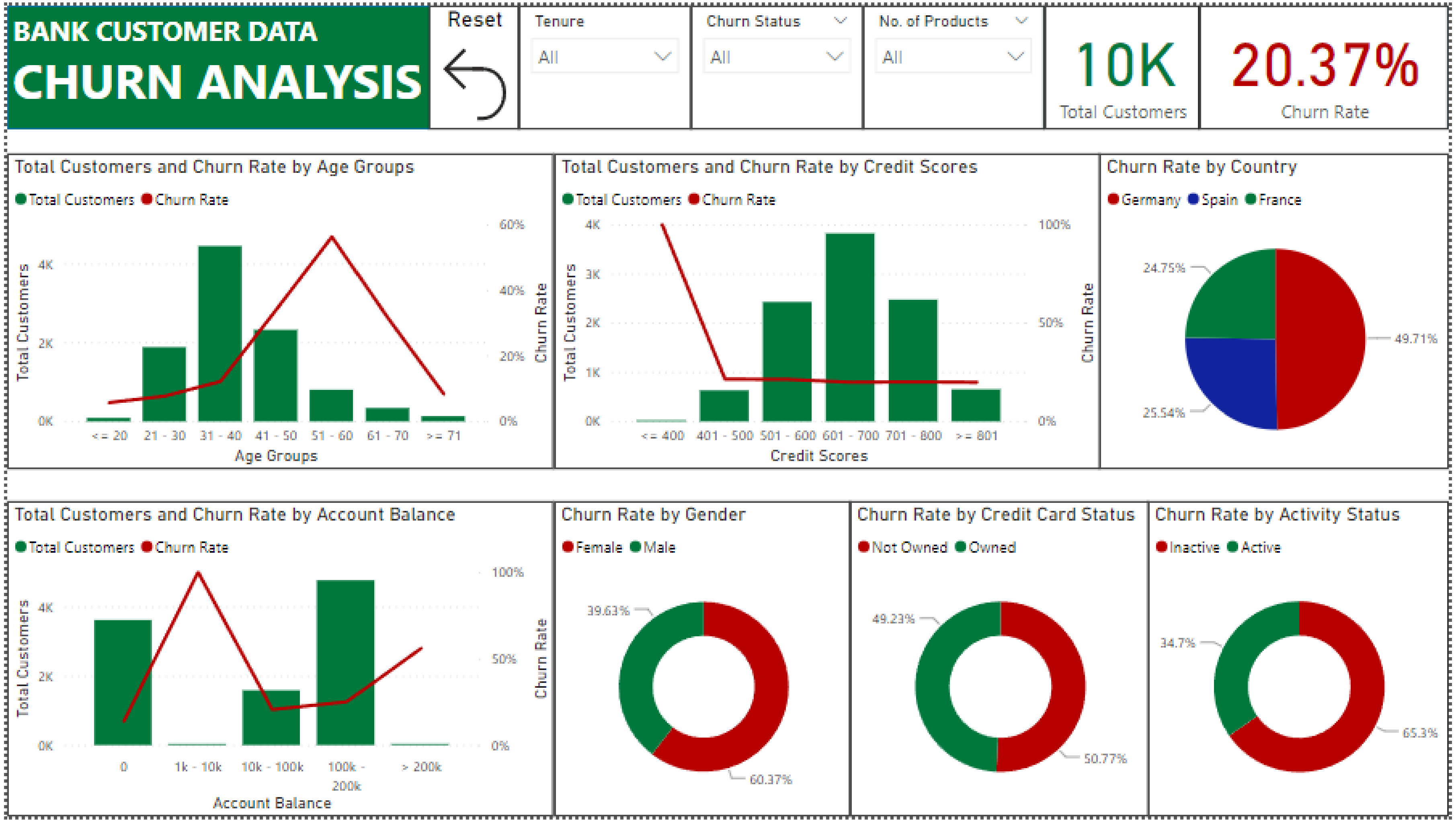
Data Wrangling:
- Removed unnecessary columns to streamline the dataset.
- Employed one-hot encoding for gender and country columns to enhance model compatibility.

Model Selection:
- Partitioned the dataset into training and testing subsets.
- Utilized GridSearchCV on logistic regression, SVM, KNN, and decision tree algorithms to identify the optimal hyperparameters.
- Selected the model that outperformed others across all evaluation metrics.

The Jupyter notebook, which includes the actual models and their evaluations, can be accessed through the following link.
https://github.com/Madhavananalyst/datascience/blob/main/Customer%20Churn%20Prediction.ipynb

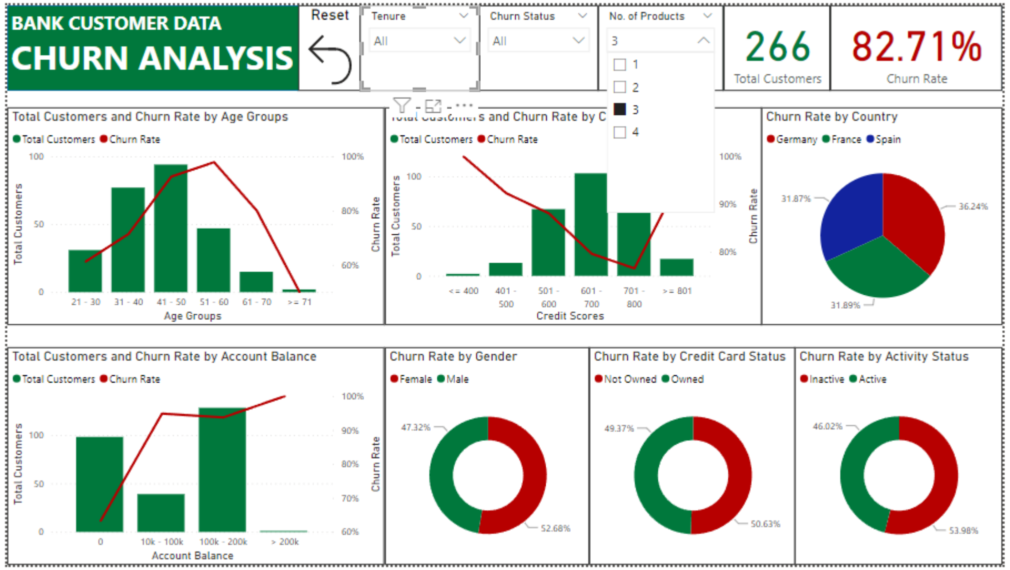
Analysis Results

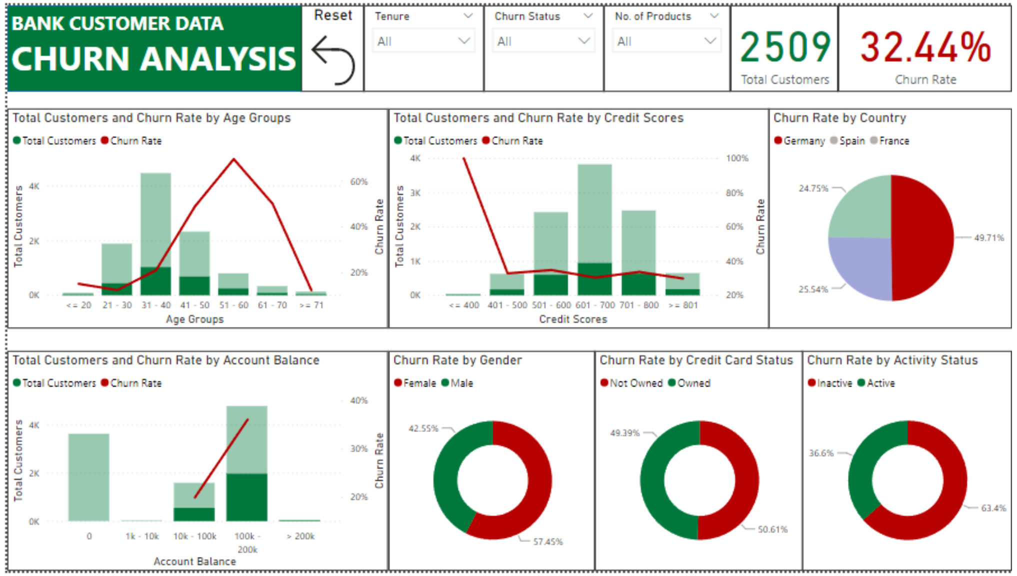
The Power BI dashboard has been developed to incorporate all the features related to churn, as outlined below.



Analysis Results

- In this Power BI dashboard, you have the option to customize your view by selecting the tenure year, churned status, and the number of products using dropdown menus.
- Additionally, the dashboard charts are interactive; by interacting directly with the charts, you can filter all the content based on the selected categories.
- If you wish to reset all the filters, simply click on the reset icon while holding down the Ctrl key.





Analysis Results

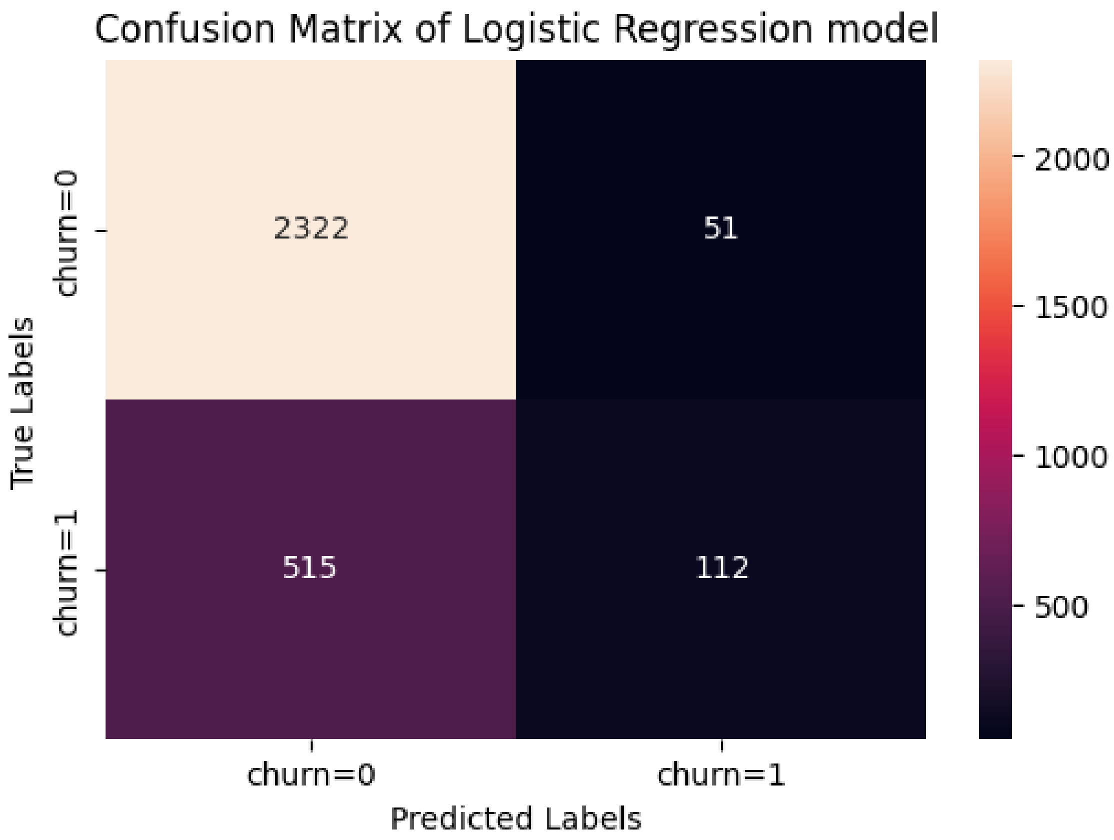
Key Insights:

- The total number of customers is 10,000, with an overall churn rate of 20.37%.
- Customers who have purchased 4 products show a 100% churn rate, while those who have purchased 3 products have an 82.71% churn rate.
- The age group of customers ranging from 51 to 60 exhibits the highest churn rate, standing at 56.21%.
- Customers with a credit score below 400 experience a 100% churn rate.
- Germany holds the highest churn rate at 49.71%.
- Customers with a bank balance ranging from 1,000 to 10,000 European dollars experience a 100% churn rate.
- The churn rate among female customers is 60.37%, indicating a higher rate compared to male customers.
- Inactive customers exhibit a higher churn rate of 65.3% compared to active customers.

Prediction Results

Logistic Regression Model:

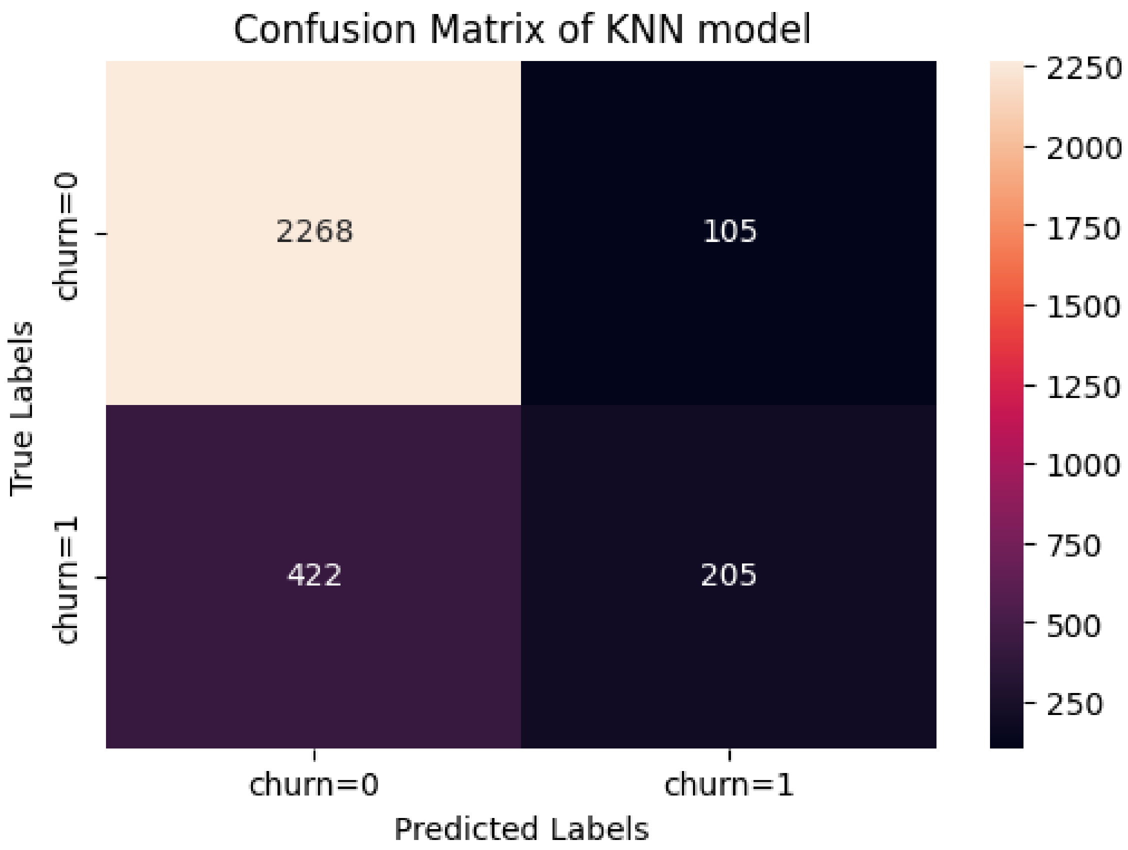
- The accuracy score is 0.8113.
- The Jaccard score is 0.8040.
- The weighted average f1 score is 0.76.
- The log loss is 0.4333.
- The confusion matrix is given.



Prediction Results

KNN Model:

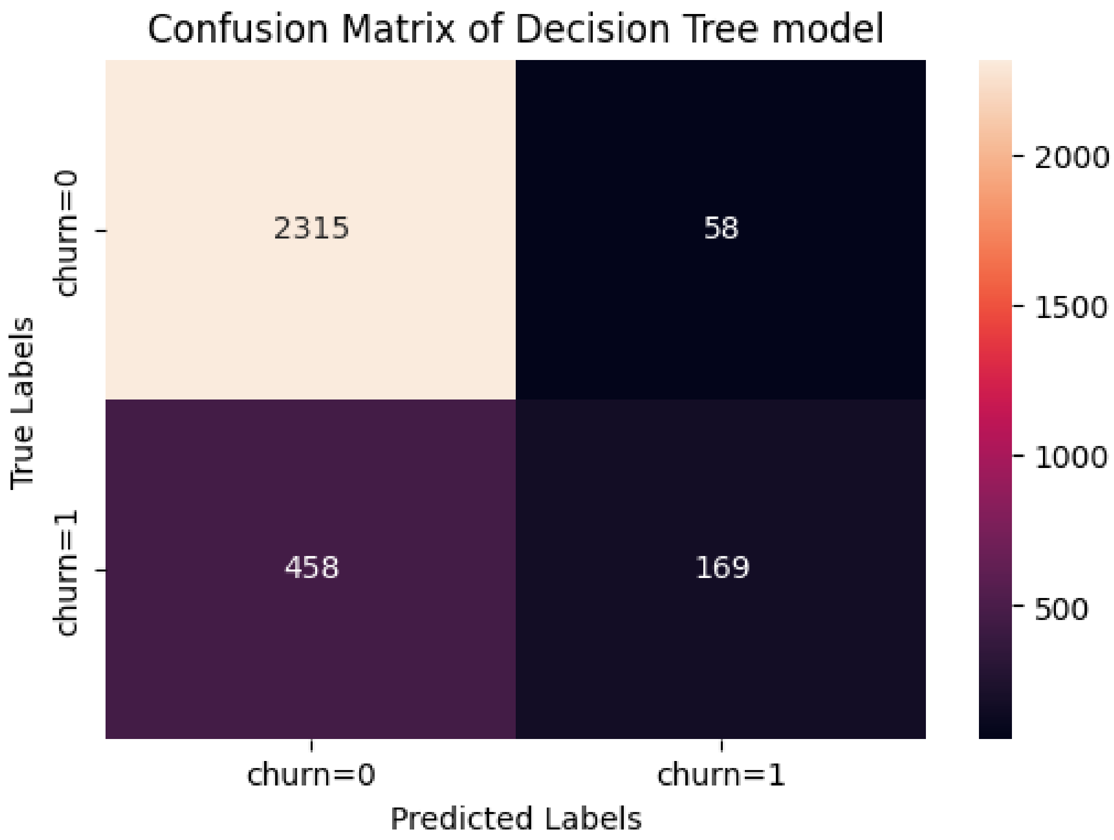
- The accuracy score is 0.8243.
- The Jaccard score is 0.8114.
- The weighted average f1 score is 0.80.
- The confusion matrix is given.



Prediction Results

Decision Tree Model:

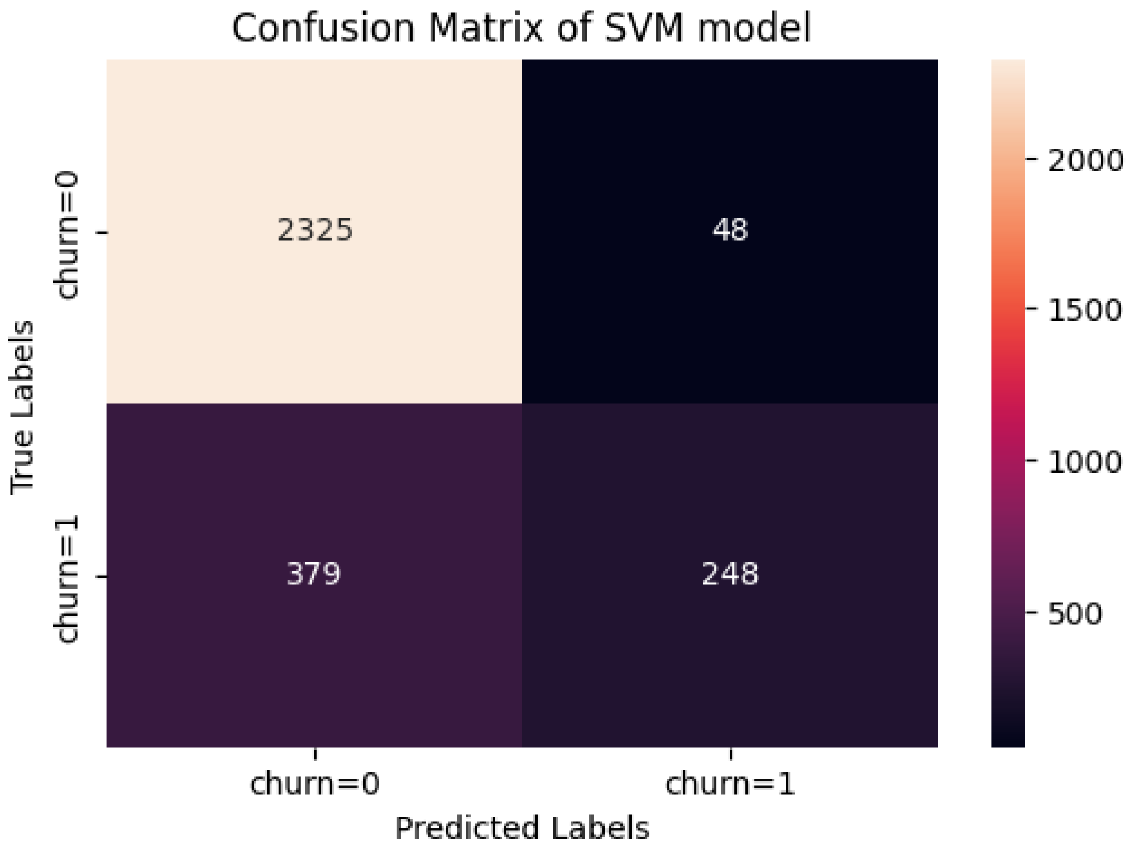
- The accuracy score is 0.828.
- The Jaccard score is 0.8177.
- The weighted average f1 score is 0.79.
- The confusion matrix is given.



Prediction Results

SVM Model:

- The accuracy score is 0.8576.
- The Jaccard score is 0.8448.
- The weighted average f1 score is 0.84.
- The confusion matrix is given.
- After evaluating various performance metrics, including the average score, Jaccard score, F1 score, and reviewing the confusion matrix, it is evident that the Support Vector Machine (SVM) model outperformed the other three models in all aspects.



Conclusion

- Our exploration into customer churn has revealed valuable insights crucial for business strategies.

- The Power BI dashboard serves as a comprehensive visual representation, showcasing the impact of various factors on churn rates.

- The prediction methodology, implemented through Jupyter notebooks, delivers actionable results by employing advanced models.

- Through these methodologies, we not only understand the current state of customer churn but also empower decision-makers with predictive tools to mitigate future attrition effectively.