Customer Churn Prediction Model

Shubhi Phartiyal

Overview

- In the competitive banking sector, retaining existing customers is a top priority.
- This project aims to predict which customers are likely to churn (close their accounts) using machine learning.
- By proactively identifying churn risks, the bank can intervene in time to reduce revenue loss.
- We followed the Data Science Process:

Problem Definition → EDA → Modeling → Evaluation → Recommendation

Understanding The Problem

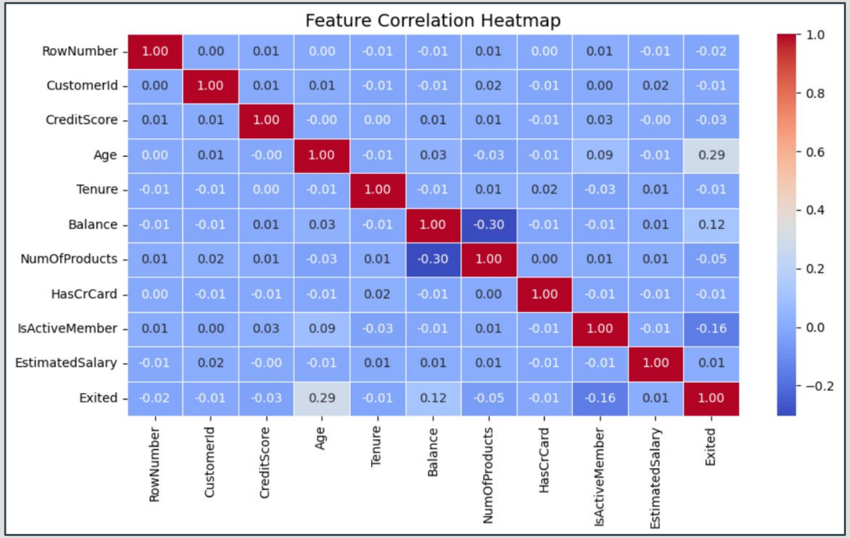
Question 1	Question 2	Question 3
Can we predict which customers are likely to churn?	What characteristics are most associated with customer churn?	How can this information help in reducing churn?

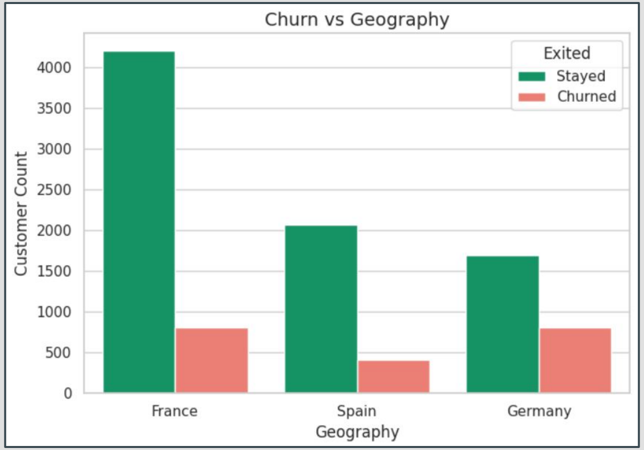
Understanding The Data:

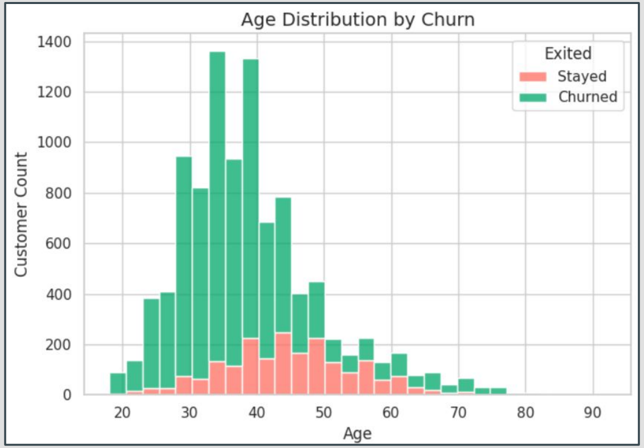
Dataset from a European bank with 10,000 customer records.

Features include:

- Credit score
- Age
- Balance
- Geography
- Gender
- Products
- Activity level
- Churn (Exited)



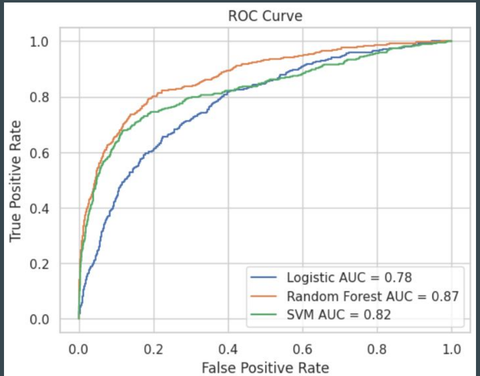




Modeling

Trained 3 classification models:

- Logistic Regression (Baseline)
- **Random Forest (Best Performance)**
- Support Vector Machine (SVM)



Random Forest Evaluation Metrics:

- Accuracy: 86%
- ROC AUC: 0.87
- Precision: 79%, Recall: 82%

Predictive Insights

Key Predictors:

- Age ↑ → higher churn
- Inactivity ↑ → higher churn
- Balance ↑ (when inactive) → higher churn
- Geography = Germany → more churn

Interpretations from the Confusion Matrix:

- TP (173): Customers correctly flagged as churn risk
- TN (1553): Customers correctly flagged as loyal
- FP (54): Bank may overinvest in customers who were not going to leave
- FN (220): Bank fails to act on real churners → high revenue loss risk

```
Confusion Matrix:
[[1553   54]
 [ 220  173]]
```

Strategic Actions to Reduce Churn

Based on the model's predictive insights, the following actions are recommended:

- Target high-risk customers with personalized retention campaigns (e.g., loyalty points, discounts).
- Re-engage inactive customers with automated email/SMS touchpoints.
- Geo-specific retention strategies, especially for customers in Germany who exhibit higher churn.
- Integrate churn scores into CRM systems to prioritize outreach based on risk level.
- Monitor key churn drivers (age, balance, activity) with internal dashboards.

Projected Business Impact

What we gain:

- 10–15% churn reduction with proactive retention measures
- Improved Customer Lifetime Value (CLV)
- Better marketing ROI through smarter targeting
- Enhanced customer experience from timely, relevant engagement

What we prevent:

- Lost revenue from undetected churners (220 false negatives)
- Wasted resources on misclassified loyal customers (54 false positives)

Answers to Business Questions:

- Yes, we can predict churn with ~87% AUC.
- Key churn drivers: Age, Balance, Inactivity, Geography.
- Insights enable cost-effective, targeted retention efforts.

The Road Ahead

Deploy → Explain → Monitor → Improve

What we accomplished:

- Built a predictive ML model to flag churn risks.
- Discovered actionable insights using EDA + feature importance.
- Linked model results to real business decisions.

Next Steps:

- Deploy model in CRM for real-time churn flagging.
- Use SHAP or LIME for explainability.
- Continuously monitor model accuracy and update with new data.

Data-driven churn prediction empowers the bank to be proactive, not reactive — keeping customers happy, loyal, and engaged.

Thank You !