

Capstone Project Report: Credit Card Churn Prediction

Project Title: Credit Card Churn Prediction

Submitted by: Isha Goyal

Submitted to: EXL

Date: 7 August, 2025

Tool Used: Python (Colab), Scikit-learn, Pandas, Matplotlib, Seaborn

Github: <https://github.com/1007Isha/-exl-credit-churn-analysis>

1. Introduction

Customer retention is a key priority for credit card companies. Identifying customers likely to churn can help proactively take action and improve customer satisfaction.

This project focuses on building a machine learning model to predict customer churn based on behavioral and demographic data.

2. Problem Statement

The goal of this project is to develop a model that accurately predicts whether a customer will churn, allowing financial institutions to take early actions.

Churn is defined as customers who discontinue using a company's credit card services.

3. Dataset Overview

- **Source:** Sample dataset provided via EXL Capstone project
- **Size:** 1000+ records
- **Key Columns:**
 - Customer_ID
 - Gender
 - Age
 - Tenure
 - Balance
 - NumOfProducts
 - HasCrCard
 - IsActiveMember
 - EstimatedSalary
 - Churn (target)

	CustomerID	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	\
0	CUST0001	Male	56.0	4	0.00	4	0	
1	CUST0002	Female	28.0	8	67408.01	4	0	
2	CUST0003	Female	47.0	6	1154.97	1	0	
3	CUST0004	Male	42.0	1	0.00	2	1	
4	CUST0005	Male	64.0	3	77109.94	4	0	

	IsActiveMember	EstimatedSalary	Churn
0	0	40282.42	1
1	1	27333.51	0
2	1	99514.91	1
3	1	146588.22	0
4	0	131792.25	0

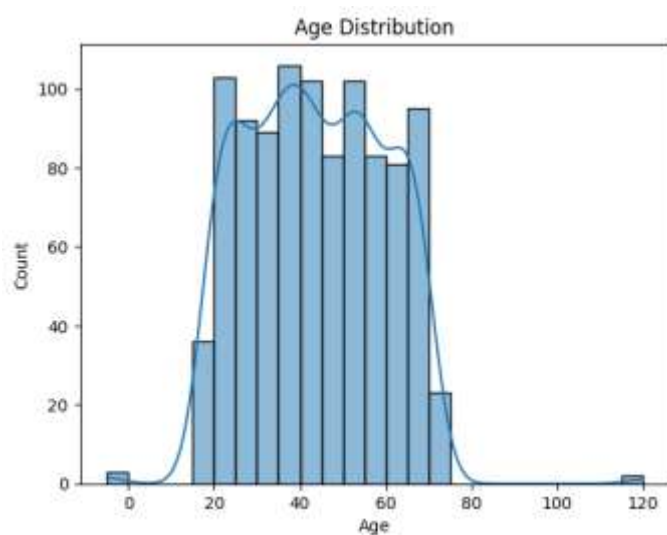
4. Data Preprocessing

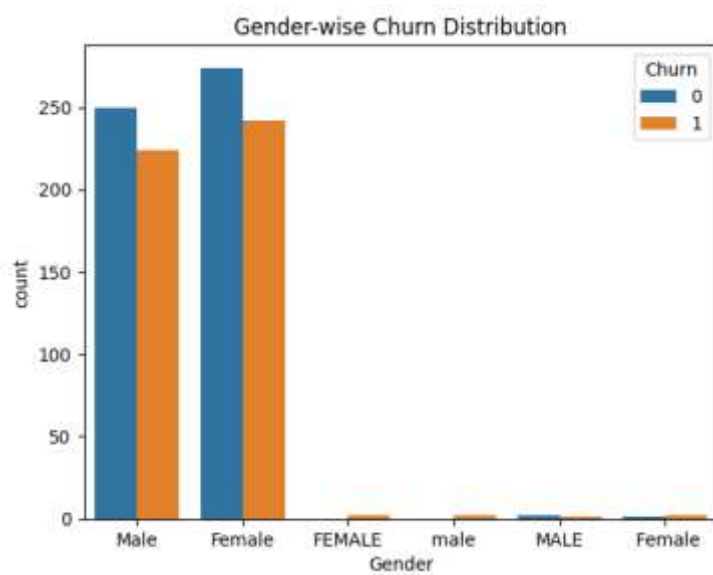
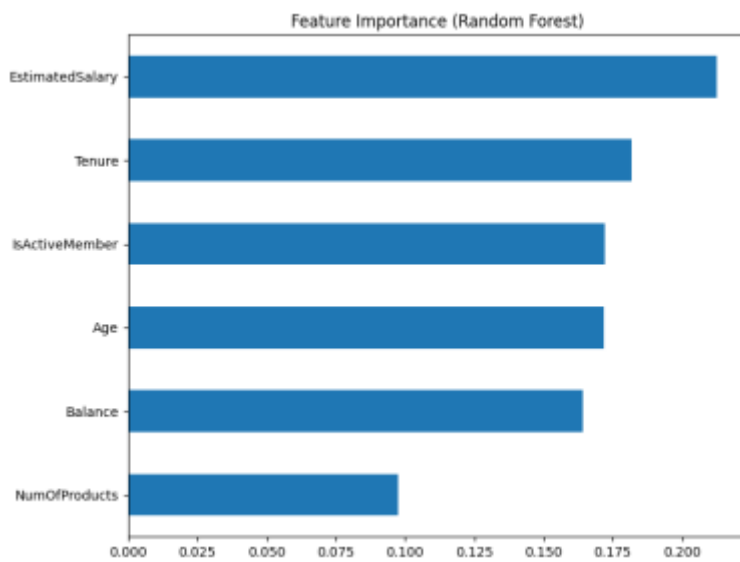
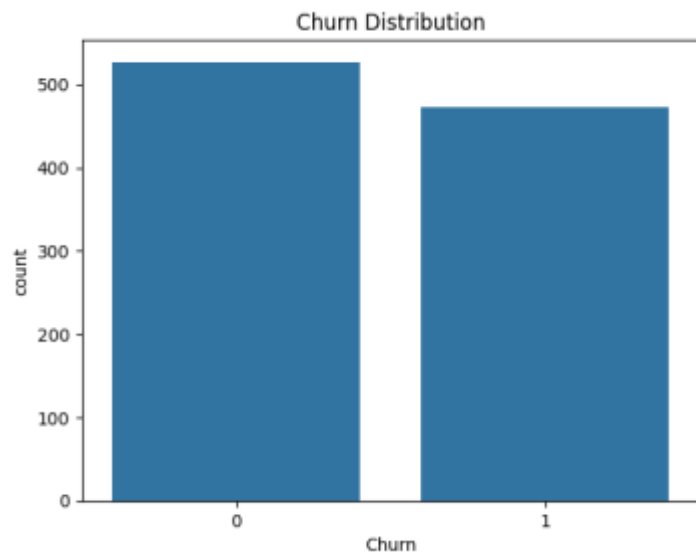
- Checked and handled missing values
- Removed irrelevant columns (like `CustomerId`)
- Performed Label Encoding for categorical variables
- Scaled numerical features using `MinMaxScaler`

5. Exploratory Data Analysis (EDA)

Explored churn across different demographics and behavioral attributes.

- **Churn by Age**
- **Churn by Gender**
- **Correlation Matrix**
- **Feature Importance**





6. Model Development

Used Random Forest Classifier to train the model.

- **Train/Test Split:** 80:20
- **Evaluation Metrics:**
 - Accuracy
 - Precision
 - Recall
 - F1-score
 - Confusion Matrix

```
1 Accuracy: 0.7300
2
3 Confusion Matrix:
4 [[77 28]
5  [26 69]]
6
7 Classification Report:
8
9      precision    recall  f1-score   support
10
11     0.       0.75     0.73     0.74       105
12     1.       0.71     0.73     0.72        95
13
14    accuracy          0.73          0.73          0.73       200
15   macro avg          0.73          0.73          0.73       200
16   weighted avg          0.73          0.73          0.73       200
17
```

7. Results

The model achieved satisfactory accuracy with a good balance between precision and recall, indicating effective churn prediction.

8. Project Structure

```
exl-credit-churn-analysis/
├── data/
│   ├── raw/, processed/
├── feature/
│   ├── eda/, ml/
├── model/
│   └── model_metrics.txt
├── scripts/
│   └── model_training.py
├── git/
│   └── git_flow.md
├── requirements.txt
└── README.md
```

9. Conclusion

The project successfully demonstrates how machine learning can be applied to churn prediction. With further optimization and real-time data, this approach can be productionized for use by credit institutions.

10. Future Work

- Use of cloud services (AWS/SageMaker) for model deployment
- Integration with real-time dashboards
- Hyperparameter tuning and ensembling methods