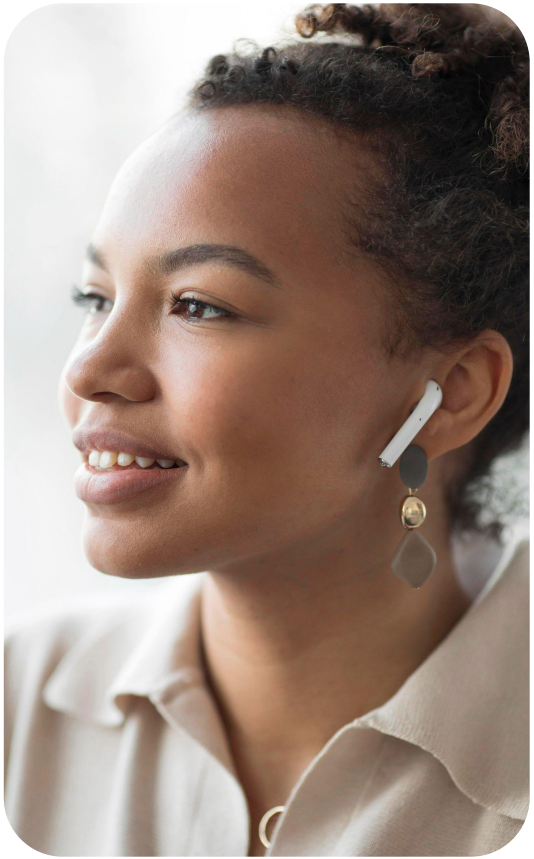
Machine Learning for Smarter Customer Insights

K-Means and Random Forest Segmentation



Unlocking Business Potential

What is Segmentation?

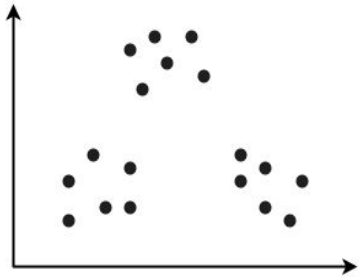
- Grouping customers by common traits to understand them better
- Helps making smarter decisions

Why automate it?

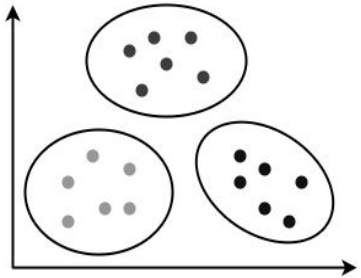
- Saves Time
- Easily Re-trained
- Scalable
- Higher Accuracy

Popular Algorithms

- K-Means Clustering (widely used)
- DBSCAN, Agglomerative Clustering and BIRCH (specific needs)



Before K-Means



After K-Means

From ML Models to Real-World Impact

1

EDA & Pre-Processing

Understand & Get Data Ready for the Models

2

K -Means

Cluster Customers with K-Means (unsupervised learning)

3

Random Forest

Classify Clusters with Random Forest (supervised learning)

4

Predictions

Cluster Labels are Predicted for New Customers

1 EDA & Pre-Processing

Dataset Description

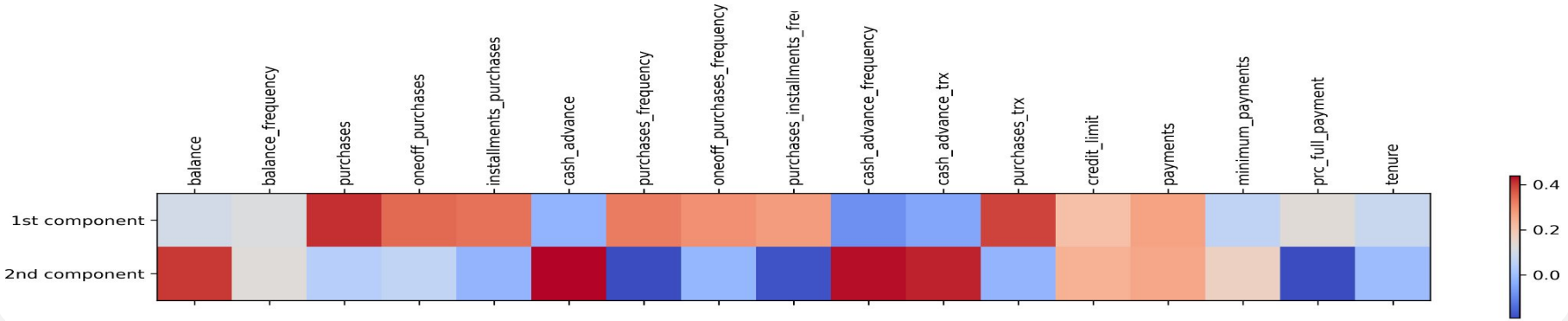
Bank Customer Dataset
9.000 Records
17 Numerical Features

EDA & Pre-Processing

Missing Values Imputation, Outlier Detection, ...
Data Distributions, Trends & Correlations

Feature Engineering

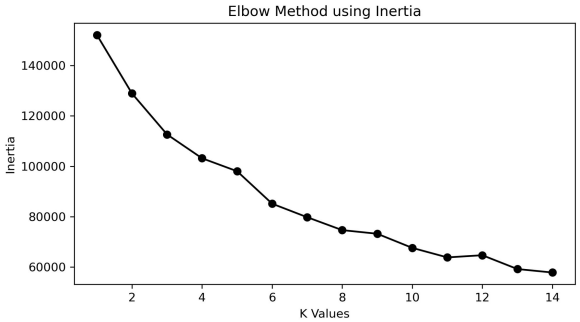
Feature Correlation
Standard Scaling
Principal Component Analysis(17D to 2 PCs)



2 K-Means

Optimal K

Elbow Method (k =4)
Silhouette Scores (0.3981)



Applying K-Means

Fitting K- Means Model to the dataset to group customers based on similarities

Cluster Analysis

Cluster Viz using PCA components, adding cluster labels to original data & Groupby for analysis

Profiling Customers

Successfully segmenting customers into 4 distinct behavioral groups

Saving Model

Saving K-Means model & clustered customer data for future use

Clustered Customers

0

Cash Advance Users

May be risky (high balance, moderate purchases and heavy cash advances)

1

Low Spenders

Might not generate much revenue (low balance, purchases and overall engagement)

2

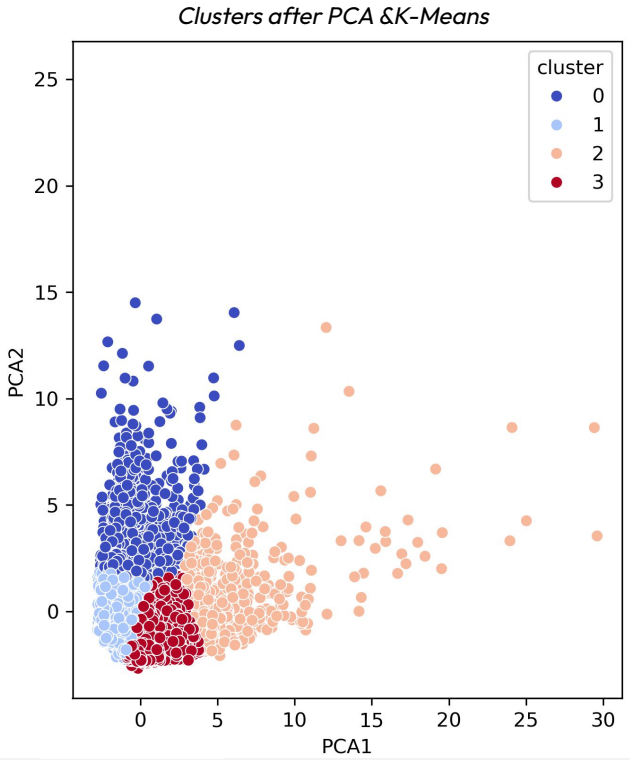
High Spenders

Most valuable segment (high balance, large purchases, high credit limits)

3

Installment Users

A a mix of good & cautious spenders (moderate balance & frequent installment)



3 Unsupervised to Supervised

1st **Model Selection**

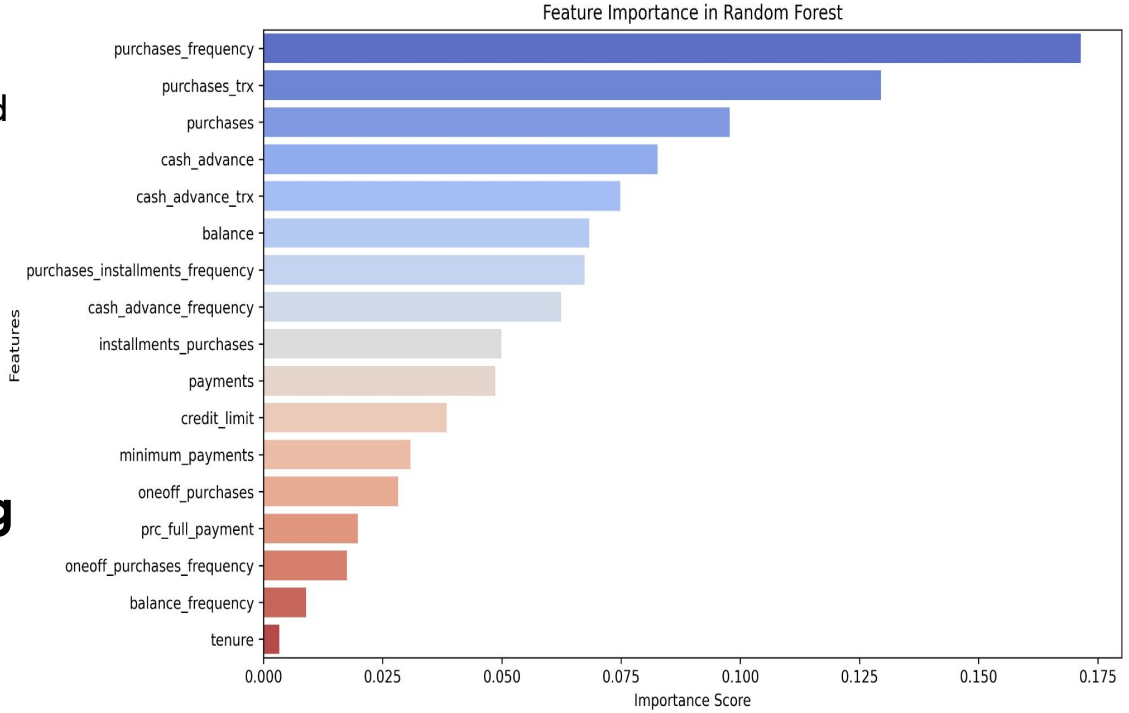
Several models were tested and evaluated (accuracy, precision, recall & F1)

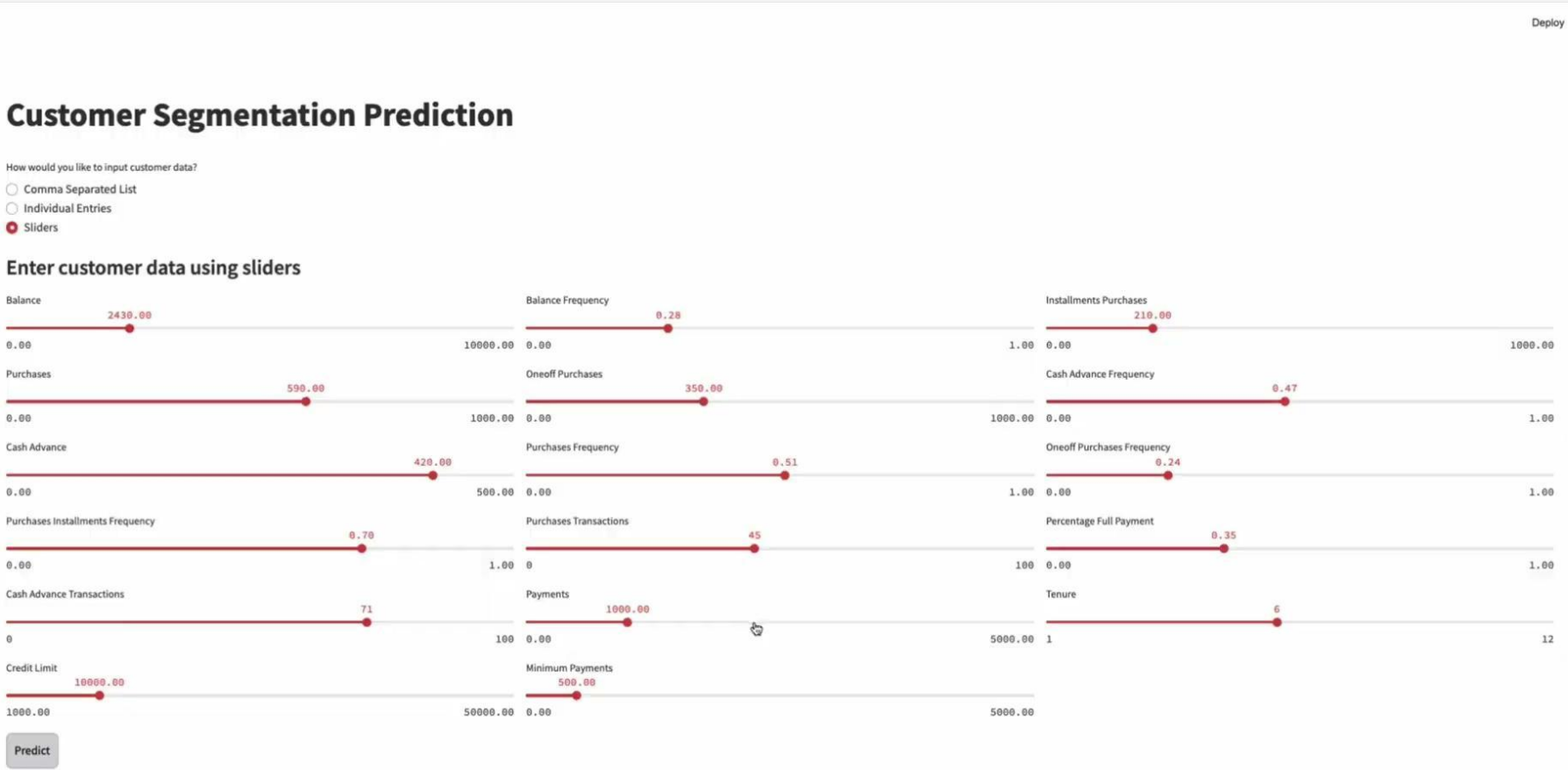
2nd **Fine-Tuning**

Didn´t significantly improved model (already well-optimized)

3rd **Analyzing & Saving**

Fitting on full data & saving best model (Random Forest) for predictions




Deploy
Customer Segmentation Prediction
How would you like to input customer data?
Comma Separated List
Individual Entries
Sliders
Enter customer data using sliders
Balance
2430.00
0.00
10000.00
Balance Frequency
0.28
0.00
1.00
Installments Purchases
210.00
0.00
1000.00
Purchases
590.00
0.00
1000.00
Oneoff Purchases
350.00
0.00
1000.00
Cash Advance Frequency
0.47
0.00
1.00
Cash Advance
420.00
0.00
500.00
Purchases Frequency
0.51
0.00
1.00
Oneoff Purchases Frequency
0.24
0.00
1.00
Purchases Installments Frequency
0.70
0.00
1.00
Purchases Transactions
45
0
100
Percentage Full Payment
0.35
0.00
1.00
Cash Advance Transactions
71
0
100
Payments
1000.00
0.00
5000.00
Tenure
6
1
12
Credit Limit
10000.00
1000.00
50000.00
Minimum Payments
500.00
0.00
5000.00
Predict

Thanks!

Do you have any Questions?