Financial Fraud Detection for Ibotta

Identify fraudulent activity in bank account transactions to <u>exclude</u> fraudsters from its platform to make new programs more effective

Objectives

Ibotta's mission is to make all purchases rewarding via digital coupons. Unfortunately, fraudsters like to take advantage of our platform.

Ibotta wants a model that takes as input one more example accounts and Outputs one or more accounts that are very similar to the supplied accounts.

Example

Fruit ID	Shape	Size	Color	Other features...	
00001234	..	..	Red	..	
0001235	..	..	Orange	..	
00012376	..	..	Red	..	
..	..	..	..	..	
..	..	..	..	..	
0012345	..	..	Orange	..	

similar fruit

Dataset

Model Pipeline

```
┌─────────────┐      ┌─────────────┐      ┌─────────────┐      ┌─────────────────┐
│ Data        │─────▶│ Data        │─────▶│ Data        │─────▶│ Feature         │
│ Extraction  │      │ Cleaning    │      │ Analysis    │      │ Extraction      │
│             │      │             │      │             │      │ & Selection     │
└─────────────┘      └─────────────┘      └─────────────┘      └─────────────────┘
MySQL,               Joining tables       Distributions etc.           │
MariaDB              etc.                                               ▼
                                                          ┌─────────────────┐
                                                          │ Train/Val/Test  │
                                                          │ split           │
                                                          └─────────────────┘
                     ┌─────────────┐                              │
                     │ Build REST  │◀─────────────────────────────▼
                     │ API using   │      ┌─────────────────┐
                     │ Flask       │◀─────│ Model Testing   │
                     └─────────────┘      │ and Selection   │
                          OR              └─────────────────┘
                     ┌─────────────┐
                     │ Make a web  │
                     │ app         │
                     │ (Stremlit,  │
                     │ Dash)       │
                     └─────────────┘
```
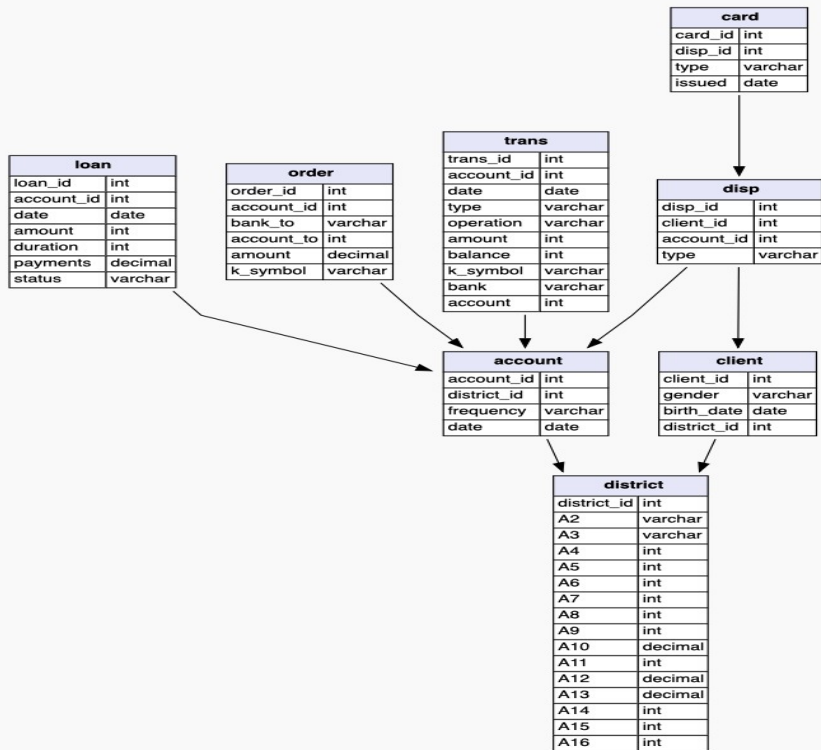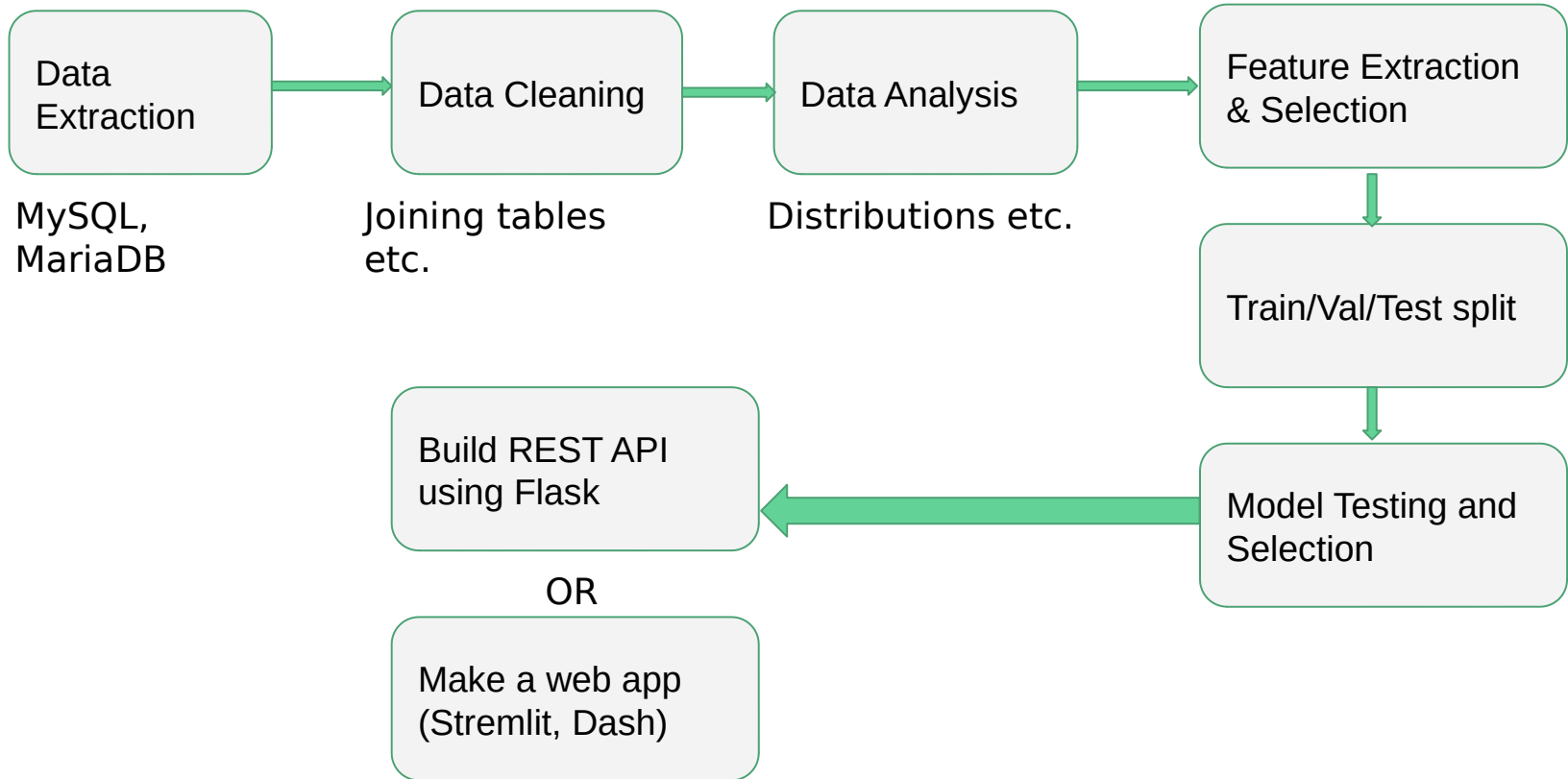
User Interface for Product Manager

Account ID under investigation
(assuming we have the entered
account ID in the database)

`0012657`

Cluster Visualization/
Outlier Detection

Summary of similar accounts as
compared to account under
investigation

Number of accounts matched	27
Confidence of accuracy of the model	80%

Feature 2

Feature 1

Select Account ID for detailed analysis

Development Timeline

Task	Start Date	End Date	Duration
Data Extraction	Jul, 26	Jul, 28	0.5 weeks
Data Cleaning	Jul, 29	Aug, 1	0.5 week
Data Analysis	Aug, 2	Aug, 4	0.5 week
Feature Extraction and Selection	Aug, 5	Aug, 12	1 weeks
Train/Test/Val Split	Aug, 13	Aug, 15	0.5 week
Model Testing and Selection	Aug, 16	Aug, 18	0.5 week
RestAPI	Aug, 19	Aug, 26	1 weeks
Web App	Aug, 26	Sep, 2	1 weeks