Sales Prediction Tool

Moetez Lahdhiri - April 2025
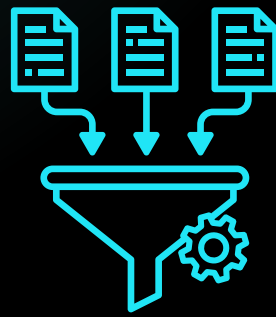
Problem Statement

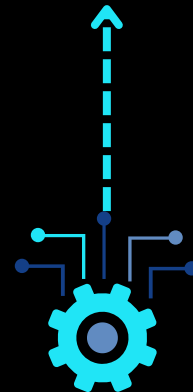| Customer ID	Product ID	Date
8449525	Opel	2017-02-03
334809	Renault	2020-01-03

Predict total transactions per customer for the next three months
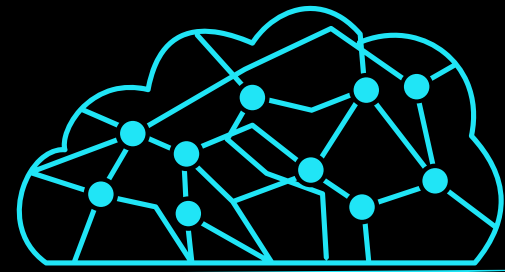
Pipelines
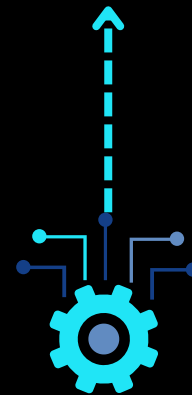
Processing Pipeline

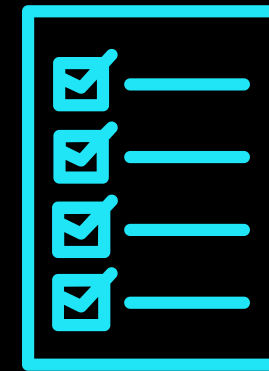Load, clean and transform data

Processing Config

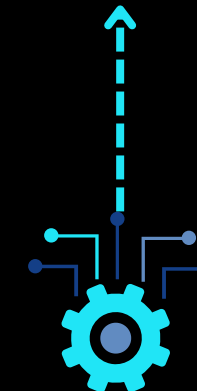### Training Pipeline

Train ML various ML models
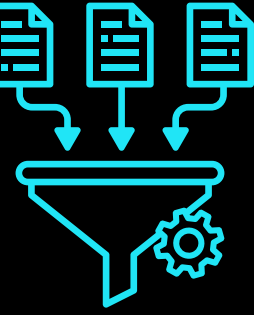
Training Config

Inference Pipeline

Perform prediction using trained models on unseen data

Inference Config

Processing Pipeline

- Level 1 : Load and clean data (remove duplicates, transform date format)
- Level 2 : Compute monthly transactions per customer & define the target variable
- Level 3 : Compute rolling features and lag features (excluding values that are not available during inference)

→ Processing config : Window size, Lag values, Number of months to predict
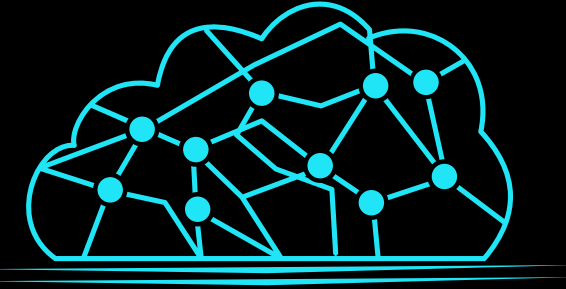
Training Pipeline

- Split the data using a backtesting approach

- Perform customer segmentation, Feature selection, Scaling

- Train various ML models

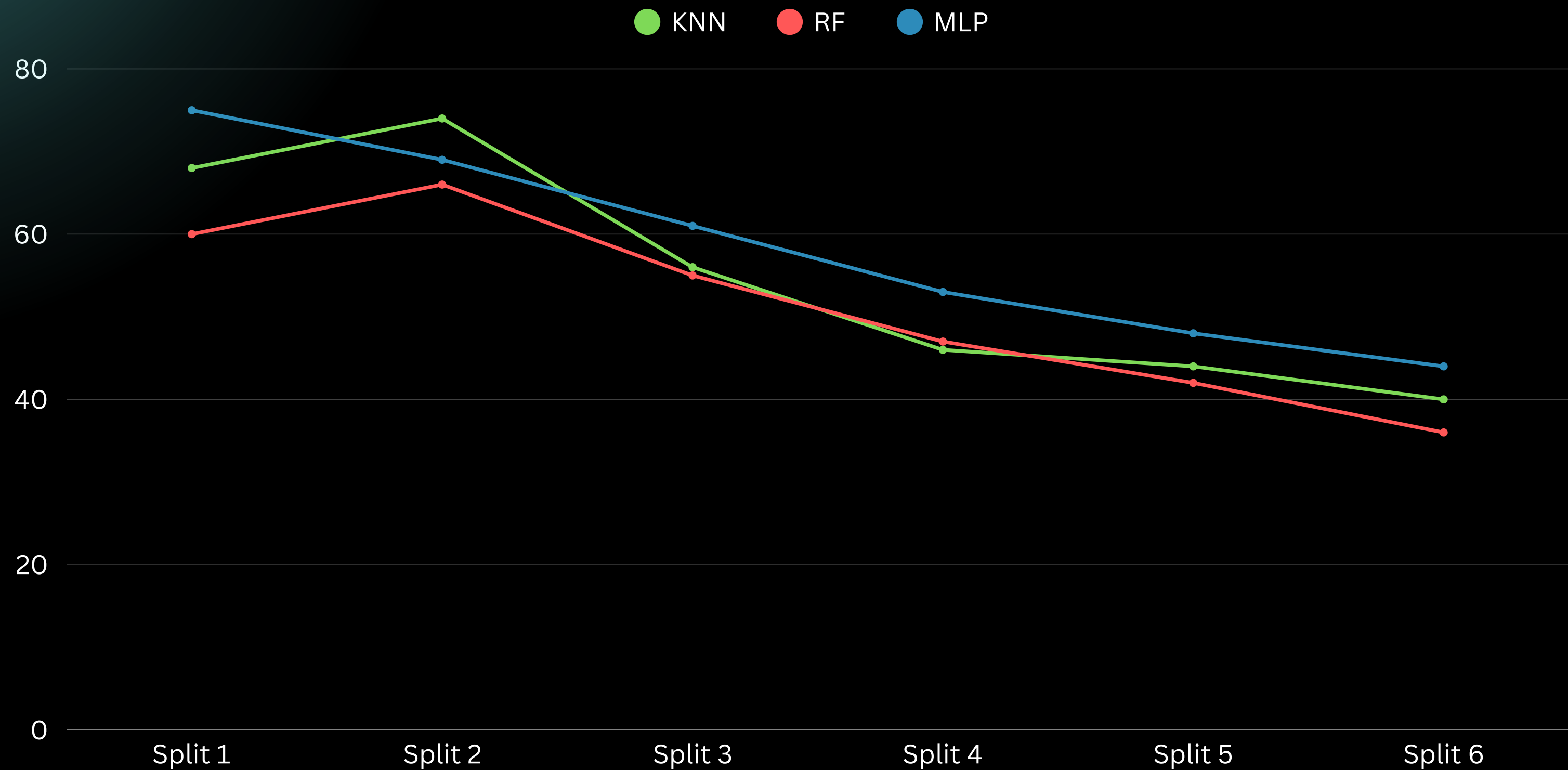- Evaluate predictions using ML mertics

- Save pipelines (for inference)

→ Training config : Splitting parameters, Segmentation parameters, ML models

parameters, features selected

Backtesting

Split	Training Period	Validation Period	Testing Period
Split 1	2017-01 to 2017-12	2018-01 to 2018-03	2018-04 to 2018-06
Split 2	2017-01 to 2018-03	2018-04 to 2018-06	2018-07 to 2018-09
Split 3	2017-01 to 2018-06	2018-07 to 2018-09	2018-10 to 2018-12
Split 4	2017-01 to 2018-09	2018-10 to 2018-12	2019-01 to 2019-03
Split 5	2017-01 to 2018-12	2019-01 to 2019-03	2019-04 to 2019-06
Split 6	2017-01 to 2019-03	2019-04 to 2019-06	2019-07 to 2019-09

Performances

Web Interface
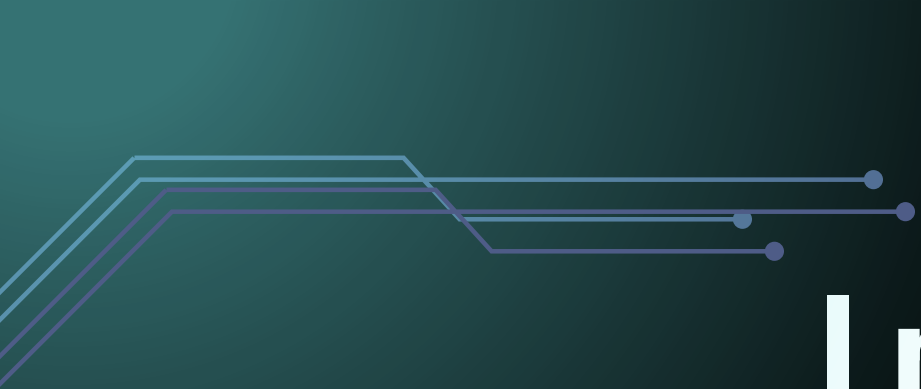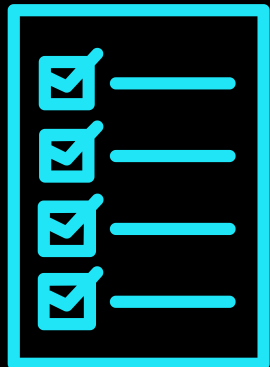
Inference Pipeline

- Load trained pipelines

- Make predictions using trained models

- Aggregate predictions

→ Inference config : Number of models to use, Aggregation strategy

Thank You