

Sales Prediction Tool

Moetez Lohdhiri - April 2025

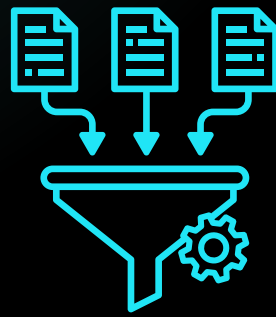


Problem Statement

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

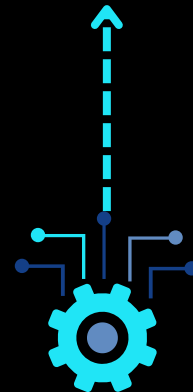
Predict total transactions per customer for the next three months

Pipelines

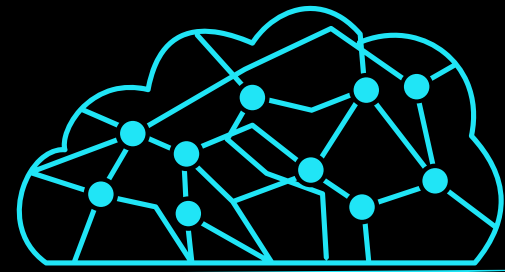


Processing Pipeline

Load, clean and transform data

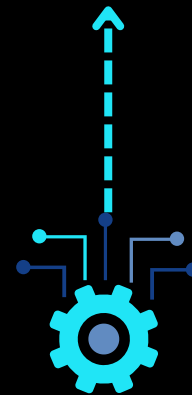


Processing Config

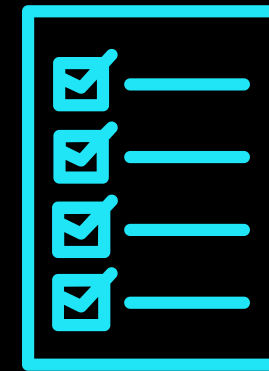


Training Pipeline

Train ML various ML models

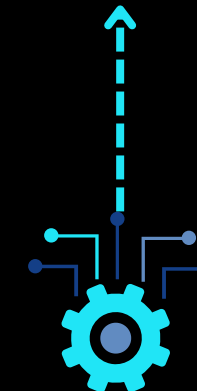


Training Config

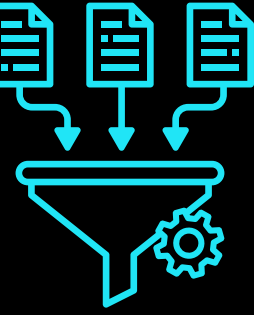


Inference Pipeline

Perform prediction using trained models on unseen data



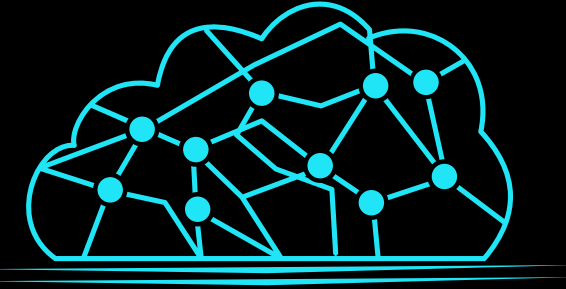
Inference Config



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 metrics
- Save pipelines (for inference)

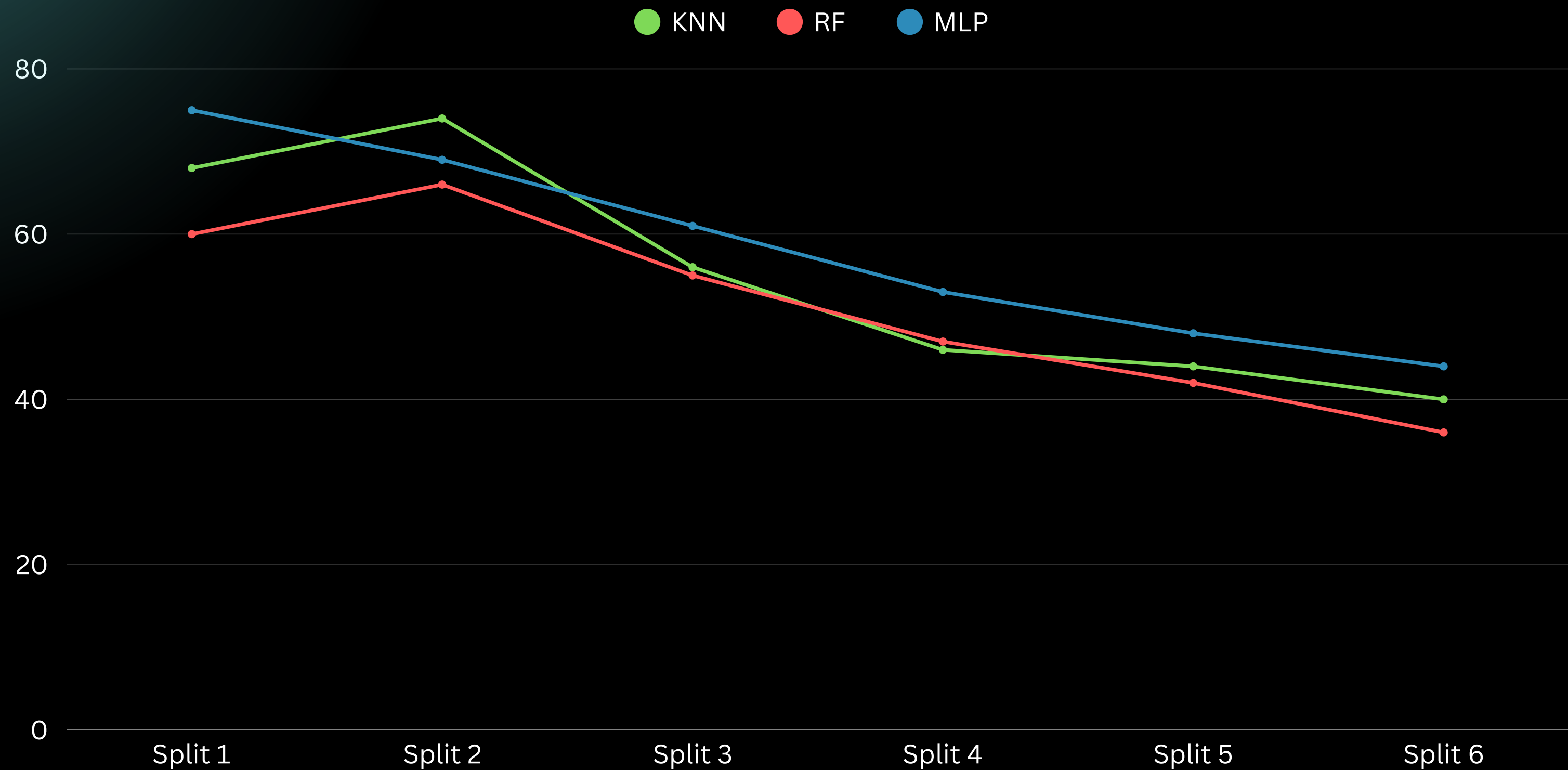
→ Training config : Splitting parameters, Segmentation parameters, ML models parameters, features selected



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

Sales Prediction - Model Training Configuration

This interface allows you to configure the parameters for the sales prediction models.

Adjust the settings below and click "Save Configuration" to save the changes. Once the pop up message is shown, click Ok then click "Run Backtesting" to run the backtesting.

Processing Configuration Parameters

Number of months to predict:

3

Minimum window size for rolling features:

12

Minimum window size for rolling features:

3

Minimum lag values to consider:

12

Minimum lag values to consider:

3

Save Processing Configuration

Training Configuration

Minimum Training Months:

12

Training Months:

3

Clustering Processor

Max Clusters:

30

Default Cluster Size:

4

ML Models

FlairNet Enabled

FlairNet Alpha

7

FlairNet L1 Ratio

0.2

KNeighborsRegressor Enabled

Number of Neighbors

10

Weights

Distance

RandomForest Enabled

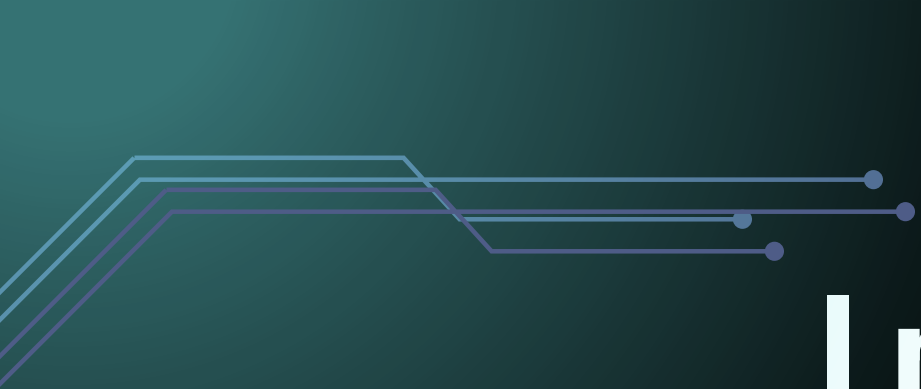
Number of Estimators

100

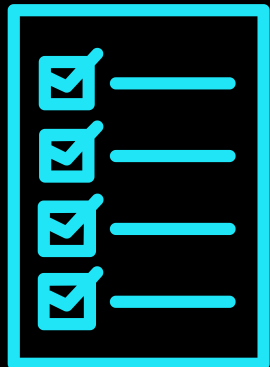
Max Depth

5

XGBoost Enabled



Inference Pipeline



- Load trained pipelines
- Make predictions using trained models
- Aggregate predictions

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

Thank You