

Submitted by: Yogya Sharma ([ys5250@nyu.edu](mailto:ys5250@nyu.edu))

## Summary Report:

This technical report provides an overview of an automated pipeline developed to collect, process, analyze, and model foreign exchange (forex) data for predictive insights. The pipeline is structured into several sequential operations, each integrating financial data services and database management systems with machine learning capabilities.

### Forex Data Collection:

Utilizing the Polygon REST Client with an API key, the pipeline fetches real-time forex data for ten currency pairs, capturing precise conversion rates. Special handling is implemented for JPY-related pairs by adjusting the scale factor (dividing by 1000).

### Database Interactions:

The pipeline interacts with MongoDB, hosted locally on port 27017. It operates on a database named "MGGYDE\_Spring24", establishing individual collections for raw forex rates and their statistical analyses. Collections are conditionally cleared and recreated to maintain data integrity for each execution cycle, ensuring no carry-over of previous state data.

### Statistical Computation:

After data insertion into MongoDB, real-time computations determine the maximum, minimum, mean rates, and volatility (calculated as the rate range over the mean) for the captured forex rates. These statistics form the basis for further technical analysis and are stored in dedicated collections with suffix "\_stats".

### Machine Learning Integration:

PyCaret's regression module streamlines the machine learning process, comparing models to identify the one with the lowest Mean Absolute Error (MAE) for the mean forex rates prediction. The best model is stored under "best\_model" and later used for predicting the mean value, where the model's performance is quantified by MAE against the latest 20 /30 data points from each currency pair.

### Synthetic Pair Synthesis:

At specific cycles (40 and 50), a synthetic forex pair is constructed by averaging the statistics of selected currency pairs. This synthesis is a form of feature engineering that inputs into the machine learning model to enhance forecasting.

### Forecastability Assessment:

The system ranks currency pairs based on MAE and classifies them into forecastability categories: 'FORECASTABLE', 'UNDEFINED', and 'NON FORECASTABLE'. This classification is saved along with the MAE values into a CSV file.

### Technical Analysis Tools:

For financial data analysis, Keltner Channel Bands and Fractal Dimensions are calculated. These technical indicators are crucial for understanding market volatility and price patterns, feeding into the fractal dimension calculation to quantify market complexity.

### Efficiency and Performance:

The code executes in cycles, representing a data collection interval. Iterations within a cycle are time-managed to synchronize with real-time data retrieval every second, with error handling for potential data fetch delays.

Below are the mean absolute error (MAE) values for the different currency pairs and their classification:

Cycle	Pair	Classification	MAE
40	GBPEUR	FORECASTABLE	0.007068103
40	GBPCHF	FORECASTABLE	0.013509158
40	GBPUSD	FORECASTABLE	0.015247341
40	EURUSD	UNDEFINED	0.031051531
40	USDCAD	UNDEFINED	0.044550131
40	EURCHF	UNDEFINED	0.055790364
40	EURCAD	UNDEFINED	0.070876003
40	USDCHF	NON FORECASTABLE	0.073261367
40	GBPCAD	NON FORECASTABLE	0.134025367
40	USDJPY	NON FORECASTABLE	0.269547682
50	GBPEUR	FORECASTABLE	0.007080661
50	GBPCHF	FORECASTABLE	0.013455659
50	GBPUSD	FORECASTABLE	0.015084714
50	EURUSD	UNDEFINED	0.031210396
50	USDCAD	UNDEFINED	0.044696811
50	EURCHF	UNDEFINED	0.055740508
50	EURCAD	UNDEFINED	0.070785385
50	USDCHF	NON FORECASTABLE	0.073067052
50	GBPCAD	NON FORECASTABLE	0.133973748
50	USDJPY	NON FORECASTABLE	0.269440787