Hochschule für Technik
und Wirtschaft Berlin
University of Applied Sciences

Fachbereich Informatik,
Kommunikation und Wirtschaft
Studiengang FAR
HTW Berlin

TITEL

Paper for the Seminar
Deep Learning

by
TOBIAS WINKE

submitted to

Fr. Dr. Alla Petukhina

February 20, 2025

Contents

1 Introduction

ANGEPASST: Predicting financial market movements has long been an area of high interest within both finance and computational research.

For predicting changes in stock prices, traditional statistical models like generalized autoregressive conditional heteroskedasticity (GARCH) and autoregressive integrated moving average (ARIMA) have been widely utilized. Still, the inherent non-linearity and volatility of financial markets frequently pose difficulties for these models [Sparks and Yurova, 2006]. Because of their capacity to identify patterns in financial time series data, data-driven techniques have grown in popularity as machine learning (ML) and deep learning (DL) have advanced. While a large proportion of research papers on this topic focus on stock market prediction, the forex market is a less frequent topic of investigation [Sezer et al., 2019]. Among these vast amounts of methods, two of them are employed in this study. One is a tree-based model, Random Forest (RF) [Breiman, 2001], and the other is a simple form of an artificial neural network (ANN), a multi-layer perceptron (MLP) [Riedmiller and Lernen, 2014], each offering unique strengths. Random Forest, an ensemble learning technique, is valued for its robustness and its ability to handle high-dimensional datasets by aggregating multiple decision trees. This method excels at feature selection and reducing overfitting. However, because RF treats each observation independently, it lacks the mechanisms needed to capture sequential dependencies inherent in time-series data. Consequently, in financial forecasting — where the temporal order of data is critical — RF may underperform relative to models like LSTM that are specifically designed to model these dependencies [Fischer and Krauss, 2018]. In this study, a rolling window framework is implemented to train and test two classes of models: (1) a multilayer perceptron (MLP) as a representative neural network and (2) random forest as an ensemble decision tree method. By evaluating the prediction metrics on out-of-sample sets, and then applying the model predictions in a backtest environment where position entries and exits are generated based on classification signals. This dual approach allows a comparison of the quality of the classification with the actual trading results.

The paper proceeds as follows. The next section provides an overview of the relevant literature and places this work in the current research on Machine Learning -based currency forecasting. Then, a description of the dataset and the preprocessing steps is given, followed by a detailed explanation of the feature engineering, the methodology of the applied models and the experimental design. Afterwards, the results of the classification algorithms and the backtesting procedure are presented. Finally, a discussion of the research and practical implications of the findings is provided and future research directions are outlined.

2 Literature Review

Pan (2024) conducted a comparative study on the effectiveness of Random Forest and LSTM in predicting gold prices. The study found that while Random Forest achieved a marginally higher predictive accuracy than LSTM, it was better suited for handling complex datasets and non-linear relationships. However, the study also identified key limitations, such as the influence of different window sizes and the lack of more complex data integration, which impacted the overall predictive ability of both models [Pan, 2024].

Wu (2024) conducted a comparative analysis of Random Forest and LSTM for stock market prediction using S&P 500 index data from 2013 to 2018. The study evaluated prediction accuracy based on Root Mean Square Error (RMSE) and scatter plots, concluding that Random Forest exhibited superior performance in a noise-free dataset. The research found that while LSTM predictions were subject to a lag effect, Random Forest provided more precise forecasts with smaller residuals. This finding aligns with previous research indicating that decision tree-based models tend to outperform deep learning models in structured and less noisy financial datasets [Wu, 2024].

Another study by Lu (2024) compared stock price prediction performance among linear models, Random Forest, and LSTM. The study found that LSTM effectively captured long-term dependencies, it exhibited slight overfitting with an RMSE of 2.42 on the training set and 1.86 on the test set. Linear regression performed well on training data but showed a decline in test performance due to stock market data's non-linearity. Notably, the Random Forest model demonstrated excellent training performance with an RMSE of 0.50 but suffered from significant overfitting, yielding a high RMSE of 33.49 on the test set. This suggests that while Random Forest is powerful in training environments, its generalization to unseen data can be problematic due to its excessive complexity and sensitivity to noise [Lu, 2024].

Moreover, Basak et al. (2019) explored the application of tree-based classifiers, specifically Random Forest and XGBoost, in predicting the direction of stock market prices.

Their study emphasized the importance of considering country-specific and industry-specific variations in stock behavior, which traditional forecasting models often fail to capture. The research introduced a classification-based paradigm shift in forecasting, achieving high accuracy in directional prediction across diverse stock markets. The study verified the robustness of the dataset using F-scores. The proposed methodology demonstrated superior accuracy (78%) compared to other techniques such as Support Vector Machines (SVM) and Artificial Neural Networks (ANN), which attained 56% and 67% accuracy, respectively. Basak et al. (2019) emphasized the importance of using ensemble methods over single decision trees, noting that the Random Forest algorithm achieved improved classification reliability [Basak et al., 2019].

This inconsistency in findings necessitates a comprehensive comparative study to understand the relative strengths and limitations of the models under different market conditions.

3 Data Description

This study utilizes historical Foreign Exchange (FX) time-series data for the EUR/USD and GBP/USD currency pairs, covering a period from 11/2008 to 11/2024 for EUR/USD and from 12/2008 to 01/2025 for GBP/USD. Major currency pairs, such as EUR/USD and GBP/USD, are among the most actively traded, which ensures high liquidity and minimizes abrupt price fluctuations that are more common in less liquid, minor pairs. The dataset consists of hourly price data in Open-High-Low-Close (OHLC) format, the date and time and the volume, obtained from a publicly accessible website [Forex Software, 2025]. Such data is typically visualized using candlestick charts, where each candle contains the OHLC information for a given timeframe. By using hourly data, this high-frequency dataset captures short-term trends and volatility patterns, which are essential for forecasting trading outcomes in the FX markets.

The raw dataset for EUR/USD includes 98338 datapoints over 6 columns and for GBP/USD there are 98314 datapoints over 6 columns.

3.1 Data Preprocessing

The raw data was collected in a CSV file and then manually cleaned to ensure its quality. Whenever the OHLC (Open, High, Low, Close) data contained errors, those entries were either removed or corrected. In particular, if an incorrect value was found in the Open or Close column, it was replaced with the corresponding value from a nearby time period—either the previous or the following hour—whenever a valid entry was available. This procedure is based on the assumption that, in a properly functioning market, the closing price at time t should closely match the opening price at time $t + 1$. Minor discrepancies were deemed negligible and were not adjusted further.

3.2 Feature Engineering

The following subsections show how the features used in the models were engineered.

3.2.1 Market Conditions

Several features have been created to display the market conditions. These are not technical indicators, but other variables whose potential influence on the models should be investigated. Most of them provide information about the current situation and positioning in the market. The binary features are encoded as booleans.

3.2.1.1 Previous Daily High

By grouping the data by days, it was possible to obtain this feature. As the name *Previous Daily High* suggests, this feature represents the highest price observed during the previous trading day. The inclusion of this feature is motivated by the idea that prior trading extremes often serve as natural resistance levels in the market.

3.2.1.2 Previous Daily Low

Similarly, *Previous Daily Low* contains information about the lowest price of the previous trading day. The inclusion of this feature is motivated by the idea that prior trading extremes often serve as natural support levels in the market.

3.2.1.3 Below Previous Daily High

Below Previous Daily High is a binary feature that provides information on whether the current price is below the highest price of the previous trading day. If the current closing price at time t is less than *Previous Daily High* at time t, then it returns the value $True$, otherwise $False$.

3.2.1.4 Below Previous Daily Low

Complementary, *Below Previous Daily Low* is also a binary feature that provides information on whether the current price is below the lowest price of the previous trading day. If the current closing price at time t is less than *Previous Daily Low* at time t, then it returns the value $True$, otherwise $False$.

3.2.1.5 Above Previous Daily High

The binary feature *Above Previous Daily High* is a counterpart to the previous features. It returns the value $True$ if the current closing price at time t is higher than *Previous Daily High*, otherwise $False$.

3.2.1.6 Above Previous Daily Low

Above Previous Daily Low is also a binary feature that returns the value $True$ if the current closing price at time t is higher than *Previous Daily Low*, otherwise $False$.

3.2.1.7 Session

The *Session* feature differentiates between major trading sessions—Asia, London, and New York—based on the hour of the day. This distinction is of interest because each session exhibits unique characteristics in terms of liquidity, volatility, and market participant behavior. For example, the London session is typically associated with higher liquidity compared to the other sessions. From 10 pm to 7 am, it returns the value 1 for the Asia session. Between 7 am and 4 pm, it returns the value 2 for the London session, and from 4 pm up to and including 9 pm, it returns the value 3 for the New York session. By incorporating the *Session* feature, the model can adjust its predictions to account for these temporal variations in market conditions, potentially leading to improved forecasting accuracy and model robustness.

3.2.2 Technical Indicators

Key technical indicators were computed to enrich the dataset with relevant market signals. These indicators include the Exponential Moving Average, the Moving Average Convergence Divergence (MACD) with the corresponding Signal Line and Histogram, the Relative Strength Index (RSI), and the Average True Range (ATR), which are formally defined as follows.

3.2.2.1 Exponential Moving Average

The Exponential Moving Average (EMA) is a weighted moving average that assigns exponentially decreasing weights to older prices, thereby allowing the indicator to respond more quickly to recent changes in the asset's price. The EMA is computed recursively as:

$$\mathrm{EMA}_n(t) = \alpha P_t + (1-\alpha)\mathrm{EMA}_n(t-1), \tag{1}$$

where P_t is the closing price at time t and α is the smoothing factor, defined as

$$\alpha = \frac{2}{n+1}. \tag{2}$$

The EMA places greater emphasis on the most recent observations, making it more responsive to price changes compared to a simple moving average. Two EMAs were employed, one uses $n = 20$ the other uses $n = 50$.

3.2.2.2 Moving Average Convergence Divergence

The MACD indicator [Appel, 2005] identifies momentum trends by computing the difference between two exponential moving averages (EMAs) of the asset price. It is defined as

$$\mathrm{MACD}(t) = \mathrm{EMA}_{\mathrm{fast}}(t) - \mathrm{EMA}_{\mathrm{slow}}(t), \tag{3}$$

where $\mathrm{EMA}_{\mathrm{fast}}(t)$ and $\mathrm{EMA}_{\mathrm{slow}}(t)$ are EMAs over different time spans, set to 12 and 26 periods, respectively.

Signal Line The Signal Line, which smooths fluctuations in the MACD, is derived by applying an EMA over a predefined window, set to nine:

$$\text{Signal}(t) = \text{EMA}_9(\text{MACD}(t)). \tag{4}$$

This component aids in identifying potential trend reversals.

MACD Histogram The MACD Histogram measures the difference between the MACD Line and the Signal Line, capturing trend strength:

$$\text{Histogram}(t) = \text{MACD}(t) - \text{Signal}(t). \tag{5}$$

A positive histogram value indicates bullish momentum, while a negative value suggests bearish movement.

3.2.2.3 Relative Strength Index

The RSI [Wilder, 1978] evaluates the magnitude of recent price changes to identify overbought or oversold conditions. It is computed as

$$\text{RSI}(t) = 100 - \frac{100}{1 + RS(t)}, \tag{6}$$

where $RS(t)$ represents the relative strength, given by

$$RS(t) = \frac{\text{Average Gain over } n \text{ periods}}{\text{Average Loss over } n \text{ periods}}. \tag{7}$$

Here, n is set to 14.

3.2.2.4 Average True Range

The Average True Range (ATR) is a measure of market volatility that considers the full range of price movement. First, the True Range (TR) is determined for each time period as the maximum of the following three values:

$$\mathrm{TR}(t) = \max\Big\{\mathrm{High}_t - \mathrm{Low}_t,\ |\mathrm{High}_t - \mathrm{Close}_{t-1}|\,,\ |\mathrm{Low}_t - \mathrm{Close}_{t-1}|\Big\}. \tag{8}$$

The ATR is then calculated as the moving average of the True Range over a specified period, n is here set to 14 periods:

$$\mathrm{ATR}(t) = \frac{1}{n}\sum_{i=0}^{n-1}\mathrm{TR}(t-i). \tag{9}$$

This indicator provides a smoothed measure of volatility and is useful for setting stop-loss levels and assessing market dynamics.

3.3 Classification Target

In this study, the target variable for classification is derived from a simulated trading strategy that incorporates dynamic stop-loss (SL) and take-profit (TP) levels based on market volatility. For each observation, a long and short position is calculated, the closing price at time t, P_t is treated as the entry price, and the ATR at time t is used to define the stop-loss and take-profit thresholds. Specifically, for a long trade, the stop-loss is set as

$$\mathrm{SL}_{\mathrm{long}} = P_t - \gamma \cdot \mathrm{ATR}_t,$$

and the take-profit is defined as

$$\mathrm{TP}_{\mathrm{long}} = P_t + \beta \cdot \mathrm{ATR}_t,$$

with $\gamma = 1.5$ and $\beta = 3.0$ as the SL respectively TP coefficient. Similarly, for a short trade the thresholds are given by

$$\mathrm{SL}_{\mathrm{short}} = P_t + \gamma \cdot \mathrm{ATR}_t, \quad \mathrm{TP}_{\mathrm{short}} = P_t - \beta \cdot \mathrm{ATR}_t.$$

The algorithm then simulates the trade over a trade horizon, meaning the subsequent H time steps (with H set to 24), to determine whether the price reaches the take-profit or stop-loss level first. The target label is assigned as follows:

$$y(t) = \begin{cases} 2, & \text{if the TP of the long position is reached before the SL,} \\ 1, & \text{if the TP of the short position is reached before the SL,} \\ 0, & \text{if neither TPs or both SLs are reached within } H \text{ steps.} \end{cases}$$

This multi-class target variable therefore captures the direction and extent of profitable moves and allows the model to distinguish between successful long and short trading signals, rather than simply classifying outcomes as profits or losses. This design provides a more comprehensive framework for evaluating model performance and improving predictive stability in the context of forex trading.

3.4 Experiment Dataset

The dataset used in this study consists of 98289 datapoints for EUR/USD and 98264 datapoints for GBP/USD, each representing a unique trading hour t for both currency pairs. Each observation includes a classification target and 15 features summarized in Table 1.

Feature	**Description**
$y[t]$	Classification Target
$x_1[t]$	Previous Daily High
$x_2[t]$	Previous Daily Low
$x_3[t]$	Below Previous Daily High
$x_4[t]$	Below Previous Daily Low
$x_5[t]$	Above Previous Daily High
$x_6[t]$	Above Previous Daily Low
$x_7[t]$	Session
$x_8[t]$	Volume
$x_9[t]$	Exponential Moving Average$_{20}$
$x_{10}[t]$	Exponential Moving Average$_{50}$
$x_{11}[t]$	MACD
$x_{12}[t]$	Signal Line (MACD)
$x_{13}[t]$	Histogram (MACD)
$x_{14}[t]$	RSI
$x_{15}[t]$	ATR

Table 1: Overview of the Dataset Features

4 Methodology

This chapter provides a formal mathematical foundation for the machine learning algorithms utilized in this study.

4.1 Random Forest

Random Forests (RF) are a robust ensemble learning technique introduced by Breiman [Breiman, 2001]. By combining multiple decision trees — each trained on a random subset of the data and features — Random Forests mitigate overfitting and typically display enhanced predictive accuracy compared to individual decision trees. Although the Random Forest framework supports both regression and classification, this section focuses primarily on its classification capabilities. Suppose a training dataset

$$\mathcal{D} = \big\{(\mathbf{x}_i, y_i)\big\}_{i=1}^{N},$$

with feature vectors $\mathbf{x}_i \in \mathbb{R}^d$ and the target variable $y_i \in \{0, 1, 2, ..., K\}$ (multi-class classification). A Random Forest classifier includes M individual decision trees, each tree trained on a bootstrap sample $\mathcal{D}_m \subseteq \mathcal{D}$. For a specific tree m,

$$\mathcal{D}_m = \big\{\big(\mathbf{x}_i^{(m)}, y_i^{(m)}\big)\big\}_{i=1}^{N},$$

where $\mathcal{D}_m$ is drawn with replacement from $\mathcal{D}$.

After obtaining the bootstrap sample $\mathcal{D}_m$, a decision tree is trained by recursively splitting the data at each node using the optimal feature I_j and threshold t_j which reduces impurity the most.

4.1.1 Stopping Criteria

To prevent overfitting and ensure computational efficiency, decision trees in a Random Forest employ stopping criteria that control tree growth. Common stopping rules include:

- Maximum Depth ($d_{\max}$): Limits the depth of the tree to avoid excessive partitioning of the data.

- Minimum Samples per Leaf ($N_{\min}$): Ensures that a leaf node contains at least $N_{\min}$ samples, preventing overly specific splits.

- Minimum Information Gain: Stops splitting if the reduction in impurity (e.g., decrease in Cross Entropy or Gini index) is below a threshold.

- Maximum Number of Nodes: Limits the total number of nodes in the tree to control complexity.

Without stopping criteria, the model would overfit and would be poor at predicting unseen test data.

4.1.2 Impurity Measure

To evaluate splits, the impurity measure Cross Entropy can be used.

Cross Entropy:

$$H(R_m) = -\sum_{k=1}^{K} \hat{p}_{mk} \log \hat{p}_{mk}$$

where $\hat{p}_{mk}$,

$$\hat{p}_{mk} = \frac{1}{N_m} \sum_{x_i \in R_m} \mathbf{1}\big(y_i \in k\big),$$

is the fraction of class k samples in region R_m.

4.1.3 Loss Function

For a given feature I_j and threshold t_j, the dataset $\mathcal{D}_m$ is split into left and right subsets, denoted as R_l and R_r. The loss function measures impurity reduction and is defined as:

$$L(I_j, t_j) = \frac{N_l}{N} H(R_l) + \frac{N_r}{N} H(R_r),$$

where N_l and N_r are the number of samples in the left and right subsets, and $H(R_l)$ and $H(R_r)$ denote their impurity values. The optimal split (I_j^*, t_j^*) minimizes this loss:

$$(I_j^*, t_j^*) = \arg\min_{I_j, t_j} L(I_j, t_j).$$

4.1.4 Prediction and Classification

For classification, each tree T_m in the forest produces an independent prediction $\hat{y}_m$. The final prediction $\hat{y}$ is determined by majority voting:

$$\hat{y} = \arg\max_k \sum_{m=1}^{M} \mathbf{1}(\hat{y}_m = k),$$

where $\mathbf{1}(\hat{y}_m = k)$ is an indicator function that counts votes for class k. Majority voting aggregates the predictions from all trees of the RF, improving stability and reducing sensitivity to noise or outliers in the training data. Since each tree is trained on a different bootstrap sample and considers a random subset of features, they offer different perspectives on the classification task. The final prediction benefits from this diversity, leading to improved generalization and a lower likelihood of overfitting compared to a single decision tree.

4.2 Multi Layer Perceptrons (MLP)

Multi Layer Perceptrons (MLP) are feedforward neural networks widely utilized for pattern recognition and function approximation in various domains, including financial forecasting [**?**, **?**]. They are suitable for examining non-linear relationships. Unlike recurrent architectures, MLP process each input sample independently, making them well-suited for scenarios where temporal dependencies have been explicitly captured or engineered through input features.

4.2.1 Architecture and Notation

A MLP is composed of three primary types of layers: an input layer, one or more hidden layers, and an output layer. The number of computation layers is referred to as the Network Depth and the number of neurons in each layer is referred to as the Layer Width. The elements in this model, despite serving different purposes, are all called neurons.

Input Layer In the input layer, the data is fed into the model. Suppose we have n input variables (features) $x_1, x_2, \ldots, x_n$, each represented as a node in the input layer. They can be highly dimensional complex objects. The number n of neurons in this layer depends on the amount of features in the model.

Weights In this fully connected model, each neuron is connected with each neuron of the neighboring layers. And each connection has an assigned weight to it. These weights are what will be adjusted in the process of training the model.

Bias The bias b_ℓ is a constant term added to the neuron's input *before* applying the activation function in the hidden layer ℓ. The bias term helps shift the activation function, thereby improving the network's ability to learn complex patterns.

Hidden Layer Within each hidden layer, every neuron h computes a weighted sum of the outputs from the previous layer $\ell - 1$ plus the bias term. Formally, let $\{h_1^{(\ell-1)}, h_2^{(\ell-1)}, \ldots, h_{j_{\ell-1}}^{(\ell-1)}\}$ be the neuron outputs of layer $\ell - 1$. For the j-th neuron in layer ℓ, the weighted sum $z_j^{(\ell)}$ is

$$z_j^{(\ell)} \;=\; \sum_{i=1}^{m_{\ell-1}} w_{ij}^{(\ell)}\, h_i^{(\ell-1)} \;+\; b_j^{(\ell)}, \tag{10}$$

where $w_{ij}^{(\ell)}$ denotes the weight from the i-th neuron in layer $\ell - 1$ to the j-th neuron in layer ℓ, and $b_j^{(\ell)}$ is the bias for the j-th neuron in layer ℓ. The activation function $\sigma(\cdot)$ is then applied to $z_j^{(\ell)}$, producing the neuron's output:

$$h_j^{(\ell)} \;=\; \sigma\big(z_j^{(\ell)}\big). \tag{11}$$

Common activation functions include the sigmoid function

$$\sigma(z) = \frac{1}{1 + e^{-z}},$$

and the Rectified Linear Unit (ReLU) [Glorot et al., 2011]

$$ReLU(z) = \max(0, z).$$

The introduction of non-linearity via the activation function enables the MLP to approximate highly non-linear functions. The output of the activation function is again sent to each neuron of the subsequent layer with its corresponding weights.

Output Layer Finally, the output layer transforms the weighted sum of the outputs from the last hidden layer into a prediction $\hat{y}$. In a multi-class classification task with three classes, the final hidden layer's outputs are converted into a probability distribution over the target classes. In a three-class classification problem, the output layer contains three neurons—one for each class. First, each neuron computes a weighted sum of the outputs from the last hidden layer. These sums are denoted as

$$z_1^{(\text{out})}, \quad z_2^{(\text{out})}, \quad z_3^{(\text{out})}.$$

These raw values are then passed through the softmax activation function, which converts them into probabilities. The softmax function is given by:

$$\hat{y}_k = \frac{e^{z_k^{(\text{out})}}}{\sum_{j=1}^{3} e^{z_j^{(\text{out})}}}, \quad \text{for } k = 1, 2, 3. \tag{12}$$

This ensures that each $\hat{y}_k$ represents the probability that the input belongs to class k, with all probabilities summing to 1. Finally, the model predicts the class corresponding to the highest probability:

$$\hat{y} = \arg\max_k \hat{y}_k. \tag{13}$$

4.2.2 Training and Loss Minimization

The MLP model is trained by comparing its predictions with the actual target values using a loss function. For classification problems, often cross-entropy loss is used, which is defined as

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{3} y_{ik} \log \left(\hat{y}_{ik} \right), \tag{14}$$

for a task with three classes, where N is the number of training samples, y_{ik} is the true label for the ith sample and class k, and $\hat{y}_{ik}$ is the predicted probability for class k obtained via the softmax activation function at the output layer.

To minimize this loss, gradient-based optimization methods are used. There are different optimizers such as stochastic gradient descent (SGD) and the Adam optimizer [Kingma and Ba, 2015], which are both commonly used and both update the model parameters (weights and biases) in the direction that reduces the loss. The training process involves the following steps:

1. **Loss Computation:** For each training sample, the model generates predictions $\hat{y}$, which are compared against the true labels y using the loss function $\mathcal{L}$.

2. **Gradient Calculation (Backpropagation):** The error, as measured by $\mathcal{L}$, is propagated backwards through the network. This process, known as backpropagation, applies the chain rule to compute the gradients of $\mathcal{L}$ with respect to each weight and bias. The gradients indicate how the model parameters should change to reduce the loss.

3. **Mini-Batch Processing:** Instead of updating the parameters after every single training sample, the data is processed in mini-batches. A mini-batch is a randomly selected subset of the training set. The gradients are averaged over the mini-batch, which results in a more stable and efficient update of the parameters.

4. **Parameter Update:** The optimizer is used to update the network parameters. It

adjusts the weights and biases based on the computed gradients and hyperparameters (e.g., learning rate). The learning rate is a hyperparameter that scales the weight updates during training, directly influencing the speed and stability with which the model converges to an optimal solution. This step is repeated iteratively over all mini-batches, thereby reducing the loss and improving model performance.

5. **Epochs:** The entire training set is processed multiple times, with each complete pass referred to as an epoch. Training continues until the loss converges or a pre-defined stopping criterion is met.

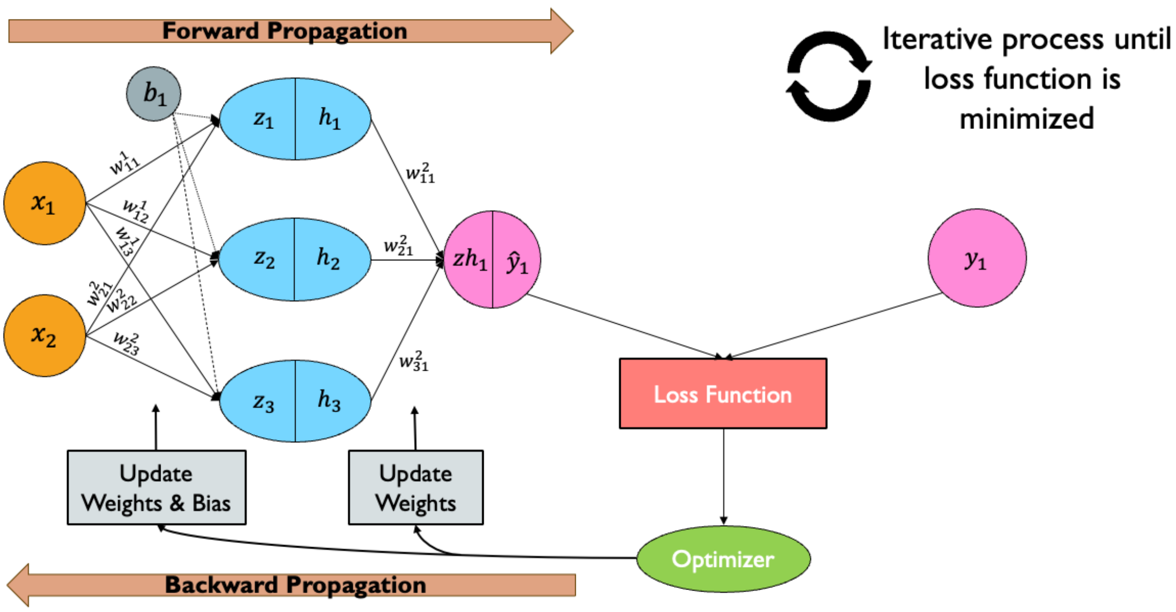

Figure 1: Exemplary illustration of a simple MLP including training (own graphic)

This structured training approach ensures that the MLP refines its parameters effectively to minimize the loss, thereby enhancing the model's prediction accuracy and overall robustness.

5 Experimental Setup

This section outlines the experiment along with the key configurations and parameters used in the preprocessing and classification stages of the experiment.

5.1 Time Rolling Window Approach for Train-Test Splitting

In this study, a time rolling windows approach is utilized to partition the time-series data into training and testing sets. Specifically, in this research, three months of historical data are used to train the model, followed by one month of data for testing. After each iteration, the window is shifted forward by one month, hence the term rolling. This procedure respects the chronological order of the data and avoids lookahead bias. Another advantage of this method is that it allows the model to adapt to non-stationarities inherent in financial markets, as each window may capture different market regimes and volatility patterns. The implementation of this approach is integrated into the Python code. 189 time rolling windows, for both currency pairs, were deployed in this study.

5.2 Trading Framework

This subsection outlines the trading frameworks used to evaluate the predictive performance of both models on FX time-series data. Two different frameworks were created; they are explained in detail later on.

Both use the same basis. First, a trading dataframe is created that includes essential market data such as the timestamp (`Gmt time`), `Open`, `High`, `Low`, `Close` prices, and the `ATR`. The Signals, provided through the model predictions for the target variable, are merged with this dataframe based on the timestamp. This unified dataset forms the basis for the backtesting. The signal for a long position is 2, the signal for a short position is 1, and 0 is the signal for no position. The stop-loss and take-profit levels of the positions are calculated using the ATR, as shown in section 3.3. The starting balance is USD 10,000 and 1% of the initial balance is risked per executed trade. Commissions and spreads are not taken into account. This means that you lose exactly USD 100 if the trade reaches the SL. The entry price of the trade is the closing price of the data point whose signal opens a trade position. The maximum trading horizon is 24 trading hours. If the trade position has reached neither TP nor SL at the end of the time horizon, the position is closed and the profit or loss at that time is realized.

Both frameworks integrate the model's predictions with market data to simulate trade execution, manage risk through SL and TP levels, construct an equity curve and compute detailed statistics for the backtest. The backtesting is performed on both pairs and for both models using both frameworks

Framework 1: "Any Signal" Backtest As the name suggests, every signal is executed in this framework. Every signal triggers a position, regardless of how many positions are already open or the fact that opposing positions are open. This framework is not reality-oriented and is very risky, as simultaneous positions on an asset are correlated with each other. If the same signal occurs several times in succession, the invested capital that is allocated to a particular price trend is stacked up, which significantly increases the risk. Opposing positions on an asset that are open at the same time on the same timeframe, over a horizon of the same length, also increase the risk, as the probability that at least one position will end in a loss consequently increases. Even though this framework is unrealistic and highly risky, it shows the raw performance of the predictive power of the models, as every prediction is utilized in this framework.

In this Framework for EUR/USD with the Random Forest predictions 26586 trades were backtested and for the MLP predictions the backtest comprises 58567 trades. For GBP/USD with the Random Forest predictions 39905 trades were backtested and for the MLP predictions the backtest contains 58044 trades.

Framework 2: "One at a time" Backtest This framework is fundamentally different from the first. As the name suggests, only one trade is executed at a time. Only when the ongoing trade has been closed by reaching the TP or SL or the maximum time horizon, can a new position be opened. All signals that occur while a position is already open are not executed. This framework is closer to reality and less risky than the other, as there is no increased risk from simultaneous or opposing positions. However, this framework cannot accurately reflect the raw model performance due to the fact that not all predictions, respectively trade signals, are implemented in actual positions.

In this Framework for EUR/USD with the Random Forest predictions XXX trades were backtested and for the MLP predictions the backtest comprises XXX trades. For GBP/USD with the Random Forest predictions XXX trades were backtested and for the MLP predictions the backtest contains XXX trades.

5.3 Python Packages

The project utilizes a variety of Python packages to facilitate data manipulation, feature engineering, model development, and visualization. Below is an overview of the key packages and their roles:

- `pandas`: Used for data manipulation and preprocessing. It provides robust data structures to handle time-series data and supports operations essential for cleaning, merging, and aggregating the dataset.
- `numpy`: Used for numerical computing, it offers support for large multi-dimensional arrays and matrices, and provides mathematical functions that are essential for feature engineering and performance evaluation.
- `scikit-learn`: This library implements various machine learning algorithms, including the ones used in the study. It also offers tools for model evaluation, such as accuracy, precision, recall and F1-score metrics.
- `matplotlib` : For plotting performance metrics and equity curves, enabling a visual evaluation of the model's behavior and the trading strategy's performance over time.

- **pandas_ta** : This package is used to compute technical indicators, used as features, from the raw time-series data.

- **Other Utilities**: Additional modules, including `copy` for deep-copying data structures and utilities from `sklearn.pipeline` and `sklearn.preprocessing.StandardScaler`, enable the MLP construction of data processing pipelines and ensure that the data is properly scaled for model training.

For model training `sklearn.ensemble.RandomForestClassifier` is used for the Random Forest model, while `sklearn.neural_network.MLPClassifier` is used for the MLP model. Hyperparameter tuning is conducted using `sklearn.model_selection.GridSearchCV` to evaluate different configurations.

The same preprocessing and classification settings are applied for both models and both pairs across all rolling time windows, maintaining uniformity in data handling.

5.4 Random Forest Configuration

For the Random Forest classifier, the hyperparameters are optimized using `sklearn.model_selection.GridSearchCV` to increase the predictive power of the model and to avoid overfitting, some hyperparameter settings were made according to subjective judgment. The model consists of an ensemble of 100 decision trees (`n_estimators = 100`). Each tree in the ensemble is trained on a subset of the data obtained by bootstrapping using Cross Entropy (`criterion = 'entropy'`) as the criterion for impurity, which measures information gain at each split to maximize separation between classes.

To control tree complexity and prevent overfitting, several stopping criteria were used; a maximum tree depth (`max_depth = 3`) is imposed, ensuring that trees remain relatively shallow and generalizable. Furthermore, to enhance stability and prevent splits based on small sample variations, the minimum number of samples required to split an internal node is set to five (`min_samples_split = 15`), and each leaf node must contain at least two samples (`min_samples_leaf = 6`). This regularization prevents overly complex trees from capturing noise in the training data. The random subset of features used to train the individual trees is three (`max_features = sqrt(n_features)`), given that our dataset includes 15 features and python applies floor rounding.

$$\texttt{max\_features} = \sqrt{15} \approx 3.87298 \approx 3$$

This means every individual tree is trained using three random features out of all 15 features. This helps to decorrelate the trees in the model.

Figure 2 shows exemplarily the architecture of the Random Forest Model. The Random Forest model is trained individually on both currency pairs and all time rolling windows, like in the Experimental Setup explained. This is important to ensure robustness across the currency pairs and different time windows. For evaluation, standard classification metrics such as accuracy, precision, recall and the F1-score are used. The backtesting statistics are also used for evaluation purposes.

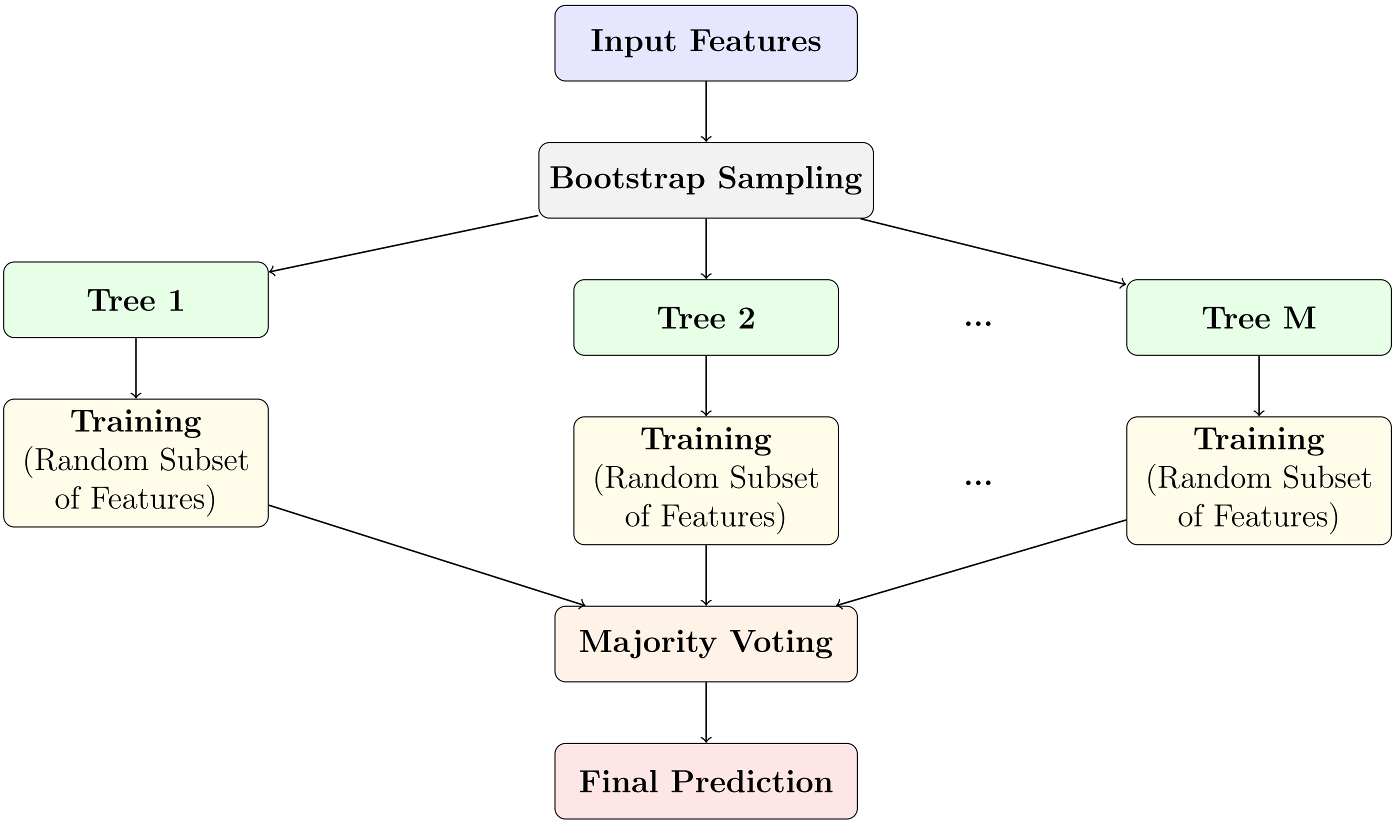

Figure 2: Random Forest Model Architecture

5.5 MLP Configuration

For the MLP model, the architecture and hyperparameters are adjusted via `sklearn.model_selection.GridSearchCV` to increase the performance, some hyperparameter settings were made according to subjective judgment. The architecture consists of two hidden layers, each with `50` neurons using a `ReLU` activation function. In the output layer the `softmax` activation function is used, given the multi-class classification task. The model is trained using the `Adam` optimizer, which is well-suited for handling non-stationary time-series data, and the `cross entropy` loss function, which is appropriate for classification tasks. A batch size of `200` is used to balance computational efficiency and gradient stability, and the model is trained for `1000` epochs to ensure adequate convergence. Figure 3 shows exemplarily the architecture of the MLP model pipeline, after training.

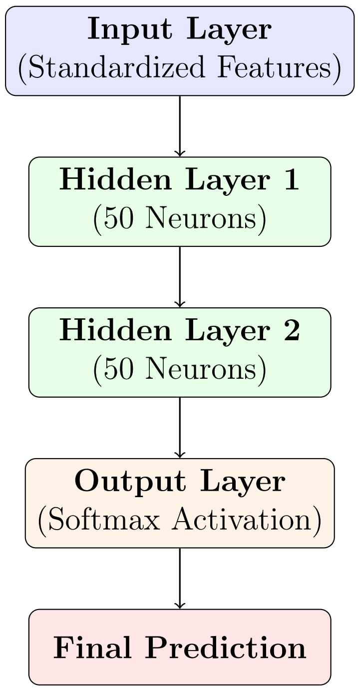

Figure 3: MLP Model Architecture

6 Results

6.1 Performance Metrics

To evaluate the classification performance of the models, standard classification metrics were employed, including accuracy, precision, recall, and F1-score. The confusion matrix, shown in Figure 4, illustrates, for a binary classification, the classification outcomes in terms of True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN). The functionality of the confusion matrix remains the same for multiclass classification. The performance metrics are computed as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{15}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{16}$$

$$F1\text{-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{17}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \tag{18}$$

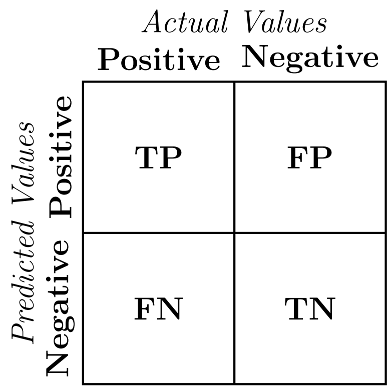

Figure 4: Confusion matrix structure.

While accuracy provides a general measure of correctness, it may not be reliable for imbalanced datasets, where one class dominates the other. In such cases, precision and recall are more informative: precision indicates how many predicted positives are actually correct, whereas recall measures how many actual positives were correctly identified. The F1-score serves as a harmonic mean of precision and recall, balancing both metrics to provide a single robust evaluation measure.

6.2 Random Forest Performance

The performance of the Random Forest models for both currency pairs and all time windows are visualized and reported in tables. Due to their length, they can be accessed in whole in the appendix, for EUR/USD in Table 10 and for GBP/USD in Table 12. However, some statistical key values, minimum, maximum, median, the 25% quantile, the 75% quantile and mean are presented here in Table 2 and Table 3 for the performance metrics.

Table 2: Summary Statistics for RF Performance for EUR/USD

	Accuracy	Precision	Recall	F1-Score
min	0.152	0.084	0.264	0.097
25%	0.359	0.182	0.332	0.218
mean	0.412	0.281	0.349	0.258
median	0.421	0.271	0.337	0.251
75%	0.468	0.352	0.368	0.297
max	0.583	0.623	0.489	0.470

Table 3: Summary Statistics for RF Performance for GBP/USD

	Accuracy	Precision	Recall	F1-Score
min	0.190	0.064	0.242	0.108
25%	0.341	0.248	0.329	0.235
mean	0.385	0.323	0.354	0.281
median	0.386	0.324	0.349	0.285
75%	0.436	0.387	0.379	0.324
max	0.576	0.693	0.467	0.465

Accuracy The accuracy values vary between 0.152 and 0.583 for EUR/USD and between 0.190 and 0.576 for GBP/USD across the time rolling windows. The evolution can be observed in Figure 5 respectively Figure 6. For EUR/USD the model has a mean of 0.412 and a median of 0.421. For GBP/USD the model has a mean of 0.385 and a median of 0.386. The small difference between the mean and median suggests the distribution is not heavily skewed, the values appear to be relatively concentrated around their central tendency. For both pairs, the accuracy is in a very similar range; it is not apparent that one of the two pairs performs better. The quantiles show that half of the values for both pairs are within a fairly small range of around 10%. Between 0.359 and 0.468 for EUR/USD and 0.341 and 0.436 for GBP/USD. The value fluctuates and has peaks above 0.50 and below 0.20, also abrupt changes in accuracy can be observed, which suggests that the models are sensitive to shifts in market regimes. This means that the classification works better in some market environments and in others the model has difficulties recognizing patterns in the data. The value never reaches a value above 0.60, which indicates that the predictive power of the model is not particularly strong, rather moderate and weak in some sections.

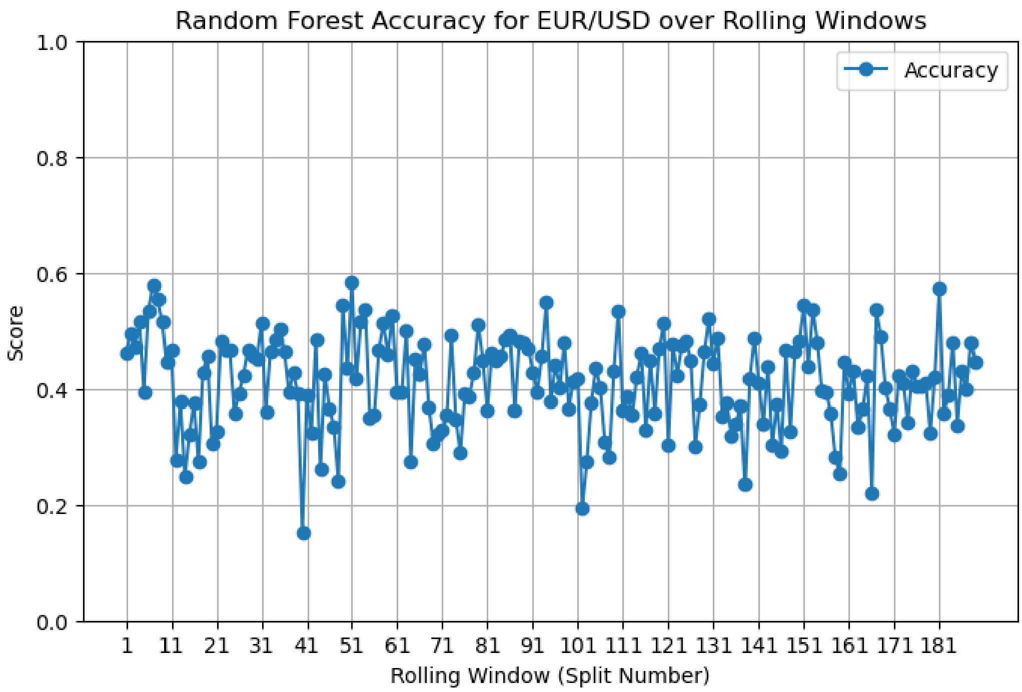

Figure 5: RF Accuracy for EUR/USD over time rolling windows

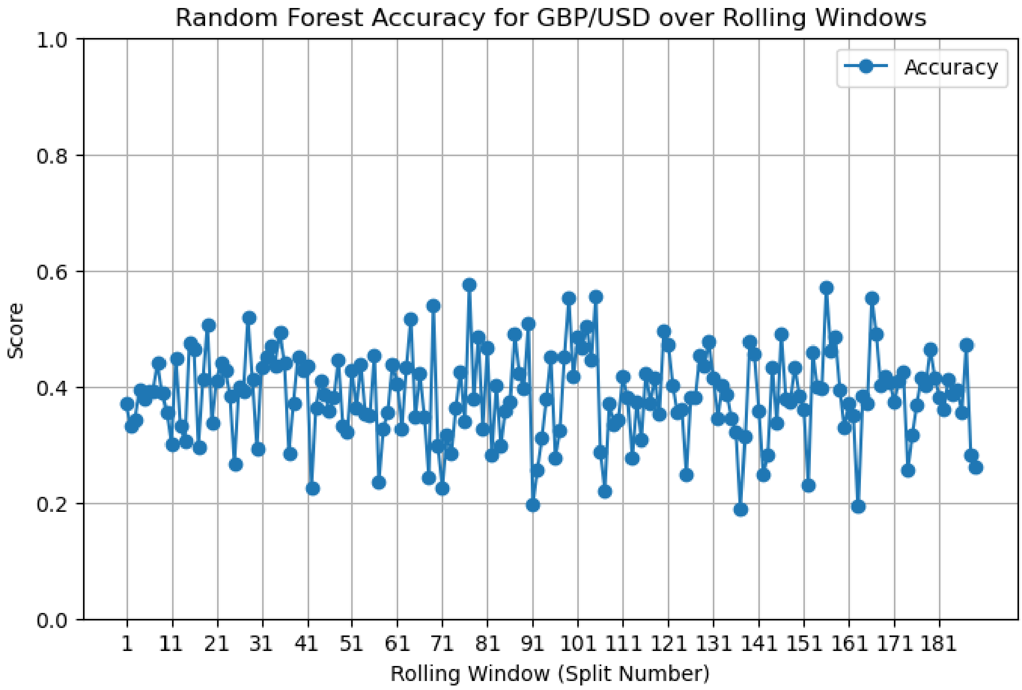

Figure 6: RF Accuracy for GBP/USD over time rolling windows

Precision The precision values range between 0.084 and 0.623 for EUR/USD and between 0.064 and 0.693 for GBP/USD across the time rolling windows. The development is shown in Figure 7 and Figure 8, respectively. For EUR/USD the model has a mean of 0.281 and a median of 0.271. For GBP/USD the model has a mean of 0.323 and a median of 0.324. Again this indicates a concentration around their central tendency. Even if GBP/USD shows a slightly wider range, the precision for both pairs is in a relatively similar range and it is not apparent that one of the two pairs performs better. Based on the rather low values for mean and median, it can be said that the RF model is not really good at predicting the correct class. A volatile dynamic can be observed in the plots, with huge leaps between different time windows. This also shows that the models struggle to consistently reflect this level of precision.

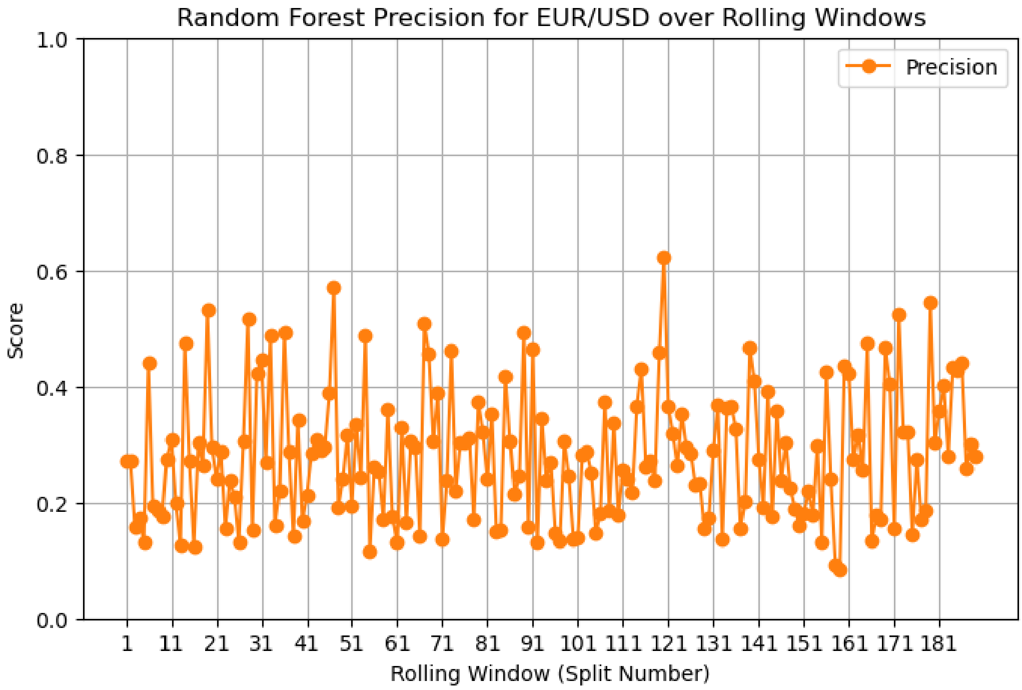

Figure 7: RF Precision for EUR/USD over time rolling windows

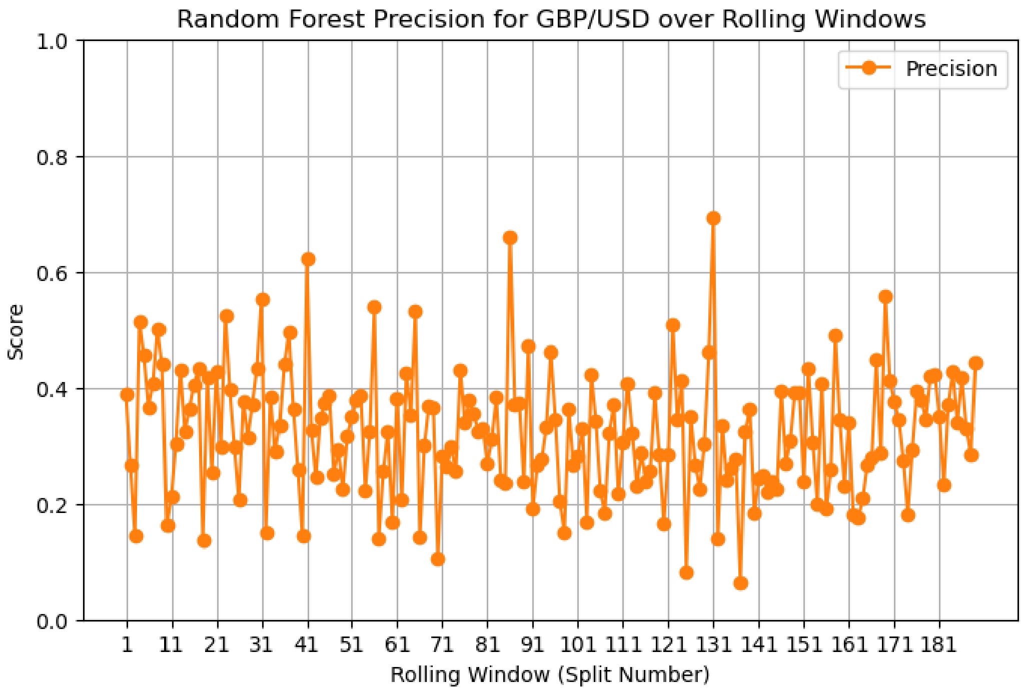

Figure 8: RF Precision for GBP/USD over time rolling windows

Recall Recall values vary between 0.264 and 0.489 for EUR/USD and between 0.242 and 0.467 for GBP/USD across the time rolling windows. The evolution can be observed in Figure 9 respectively Figure 10. For EUR/USD the model has a mean of 0.349 and a median of 0.337. For GBP/USD the model has a mean of 0.354 and a median of 0.349. This small difference aligns again with the previous metrics. Based on the quantiles it can be observed, that half of the values are in small range from 33% to 38%. In addition to the narrower statistical range the values for recall are more robust over the time windows than the previous metrics. Still the results show that the models often miss a notable fraction of the target classes, means that in most cases over 60% of the correct class is not identified.

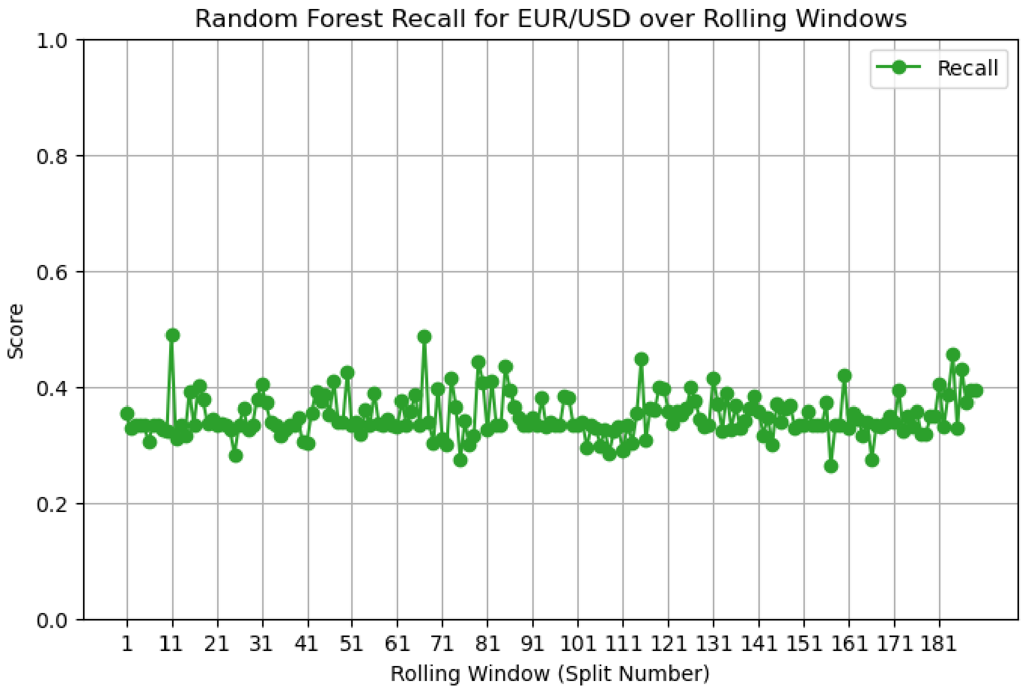

Figure 9: RF Recall for EUR/USD over time rolling windows

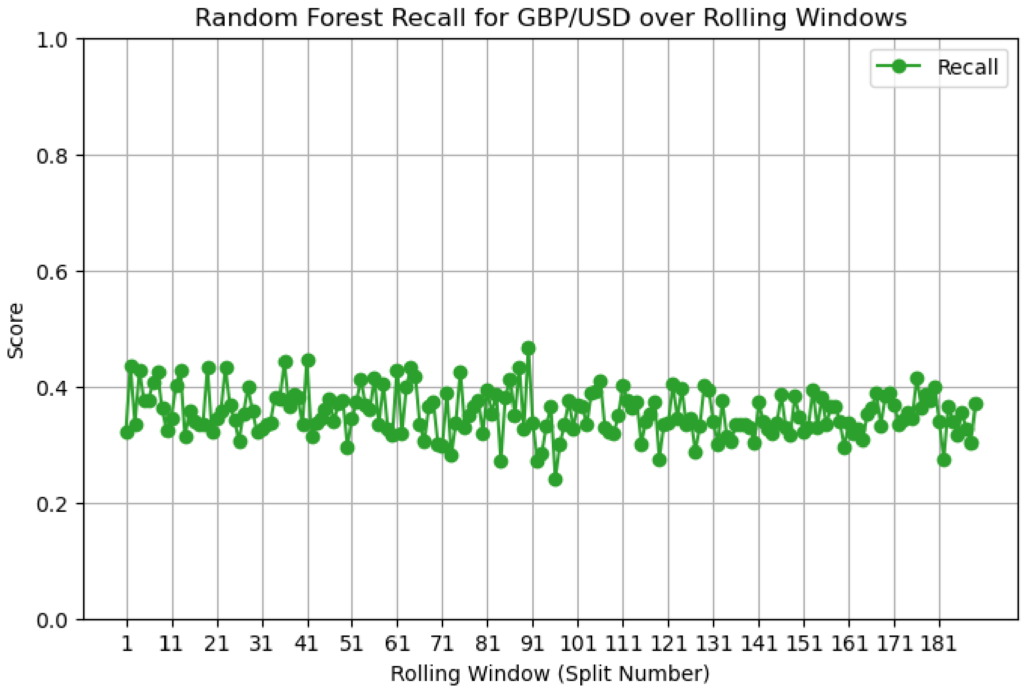

Figure 10: RF Recall for GBP/USD over time rolling windows

F1-Score The F1-scores exhibit a range from 0.097 to 0.470 for EUR/USD, with a mean of 0.258 and a median of 0.251, and from 0.108 to 0.465 for GBP/USD, with a mean of 0.281 and a median of 0.285. As visualized in Figure 11 and Figure 12, the majority of F1-scores lie between roughly 0.20 and 0.35, with only occasional peaks reaching above 0.40. This distribution aligns with the observed behavior of precision and recall, reflecting the challenge of maintaining both a high fraction of correct class predictions (precision) and capturing a significant portion of the true class (recall). The relatively low upper bounds further underscore the model's limited capacity to balance these two metrics simultaneously. Although both currency pairs exhibit similar F1-scores, GBP/USD shows slightly higher mean and median values, indicating marginally better alignment between precision and recall in its rolling windows. Overall, the fluctuations and modest levels of F1-scores confirm that the Random Forest often struggles to maintain consistent performance when confronted with shifting market regimes and changing class distributions.

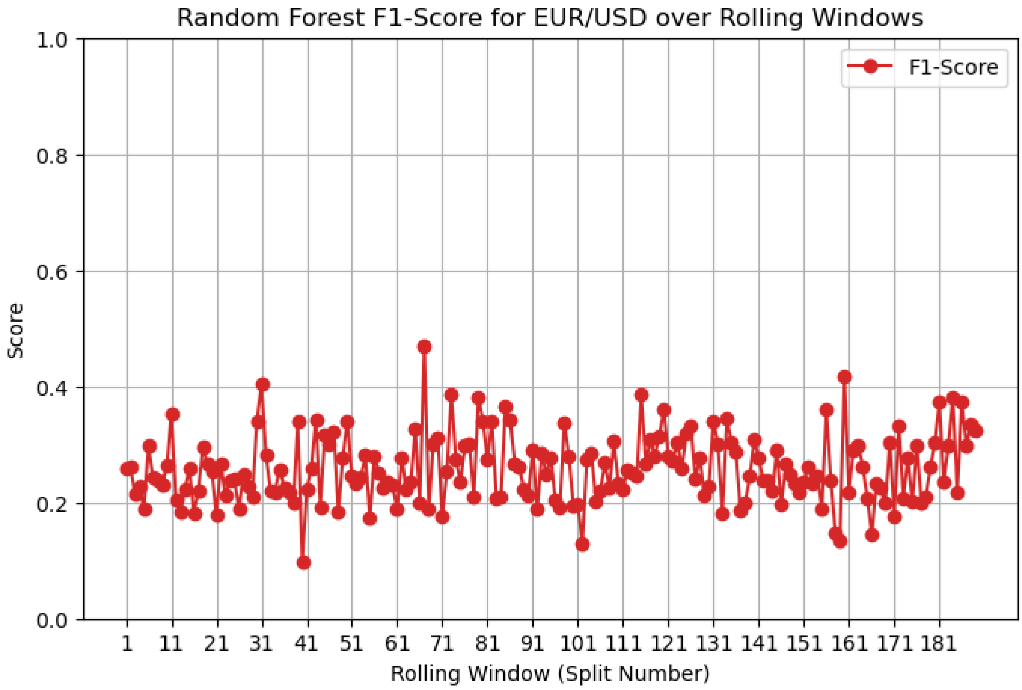

Figure 11: RF F1-Score for EUR/USD over time rolling windows

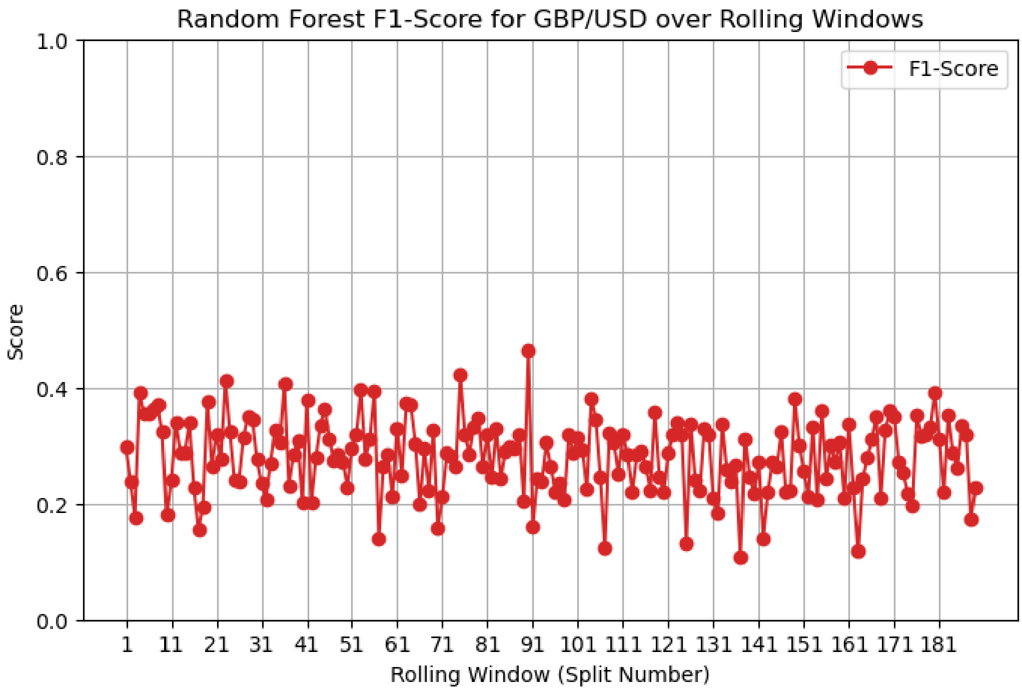

Figure 12: RF F1-Score for GBP/USD over time rolling windows

In general, the evaluation of the Random Forest model across the time rolling windows reveals moderate predictive performance with notable variability. The fluctuations across time windows underscore its sensitivity to changing market conditions and class imbalances. The similar central tendencies observed for both EUR/USD and GBP/USD suggest that, despite minor differences, the model's performance is generally consistent across these major currency pairs, but not robust over all time windows.

RF Backtesting: Framework 1 The Backtesting for Framework 1 was performed as in chapter 5.2 explained. The performance for balance values under zero were also captured, even though under real circumstances, this would not be possible like this.

Description and Interpretation Both currency pairs exhibit large negative total returns, with EUR/USD at -278.86% and GBP/USD at -563.67% (-56,367.03 $). This highlights the inherent risk of executing every signal generated by the model, including correlated or contradictory positions. The key metrics are shown in Table 4. The average yearly return is negative for both pairs in the double digits. If the backtest would have ended if the account hits zero, than for EUR/USD it would have ended 22.12.2009 and for GBP/USD 30.06.2009.

Table 4: Random Forest Backtesting Results under Framework 1

Metric	**EUR/USD**	**GBP/USD**
Total Return (%)	-278.86	-563.67
Total Return ($)	-27885.67	-56367.03
Avg Return per Year (%)	-17.43	-35.23
Number of Trades	26586	39905
Long Trades	12078	16508
Short Trades	14508	23397
Long Win Percentage (%)	34.66	35.40
Short Win Percentage (%)	36.21	35.12
Overall Win Percentage (%)	35.51	35.24
Date Account Reached 0	2009-12-22	2009-06-30

We can see that for GBP/USD 39905 trades, rougly 50% more trades than for EUR/USD

(26586), were executed. The model opened slightly fewer long trades than short trades for each pair, but neither side exhibits a strong performance advantage.Both pairs show a similar overall win rate around 35%, indicating the model struggles to produce consistently profitable trades.

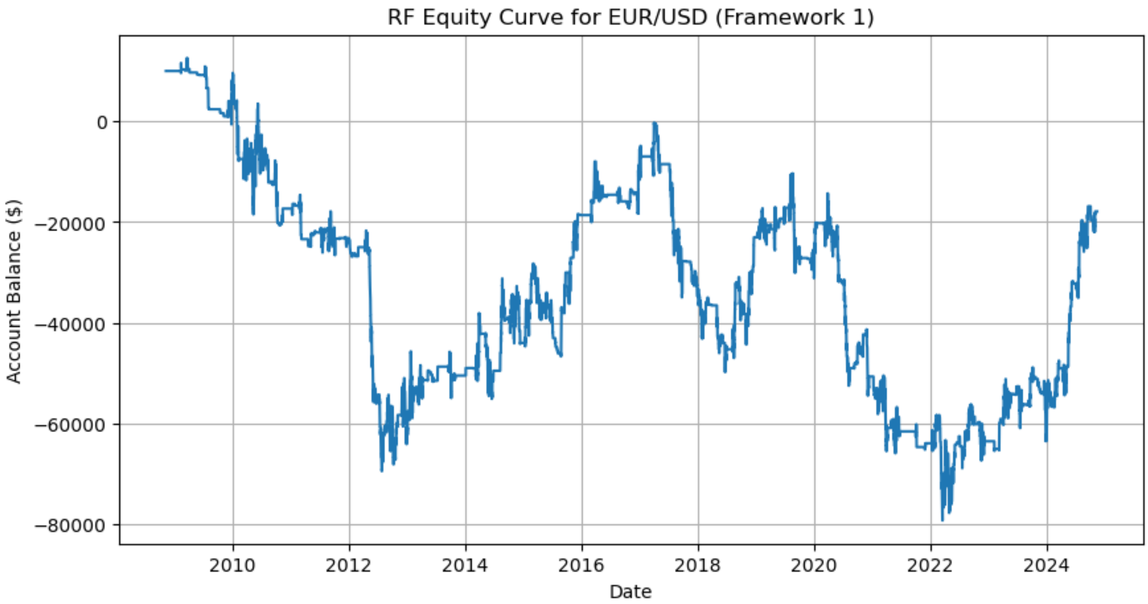

Figure 13: RF EUR/USD Equity Curve (Framework 1)

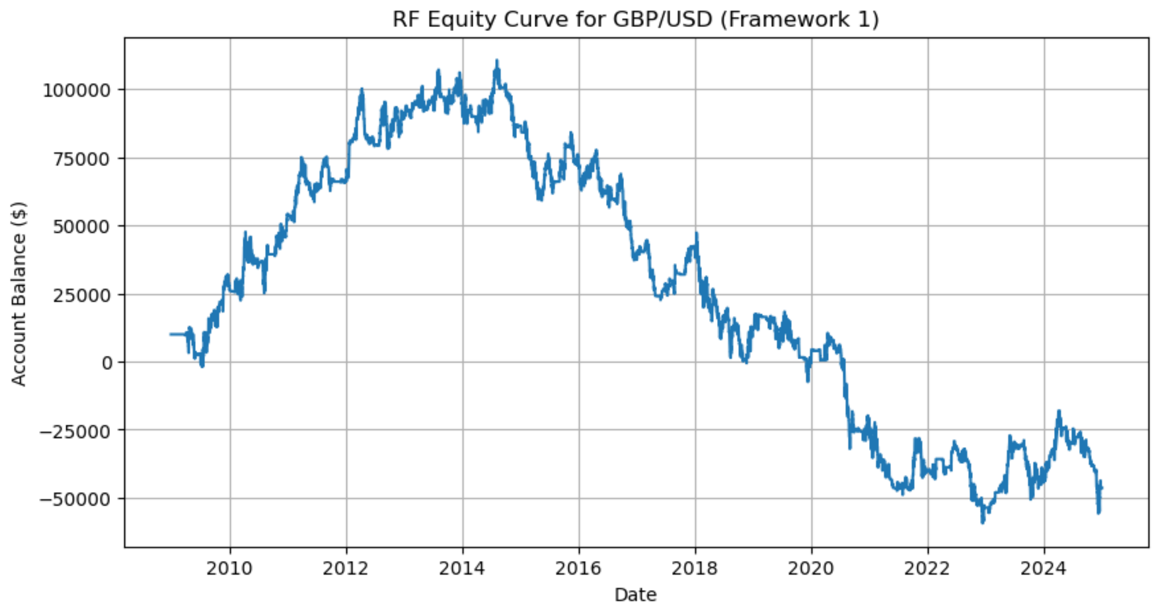

Figure 14: RF GBP/USD Equity Curve (Framework 1)

The accumulation of positions magnifies losses if market conditions move counter to the predicted trend. This framework, by design, does not incorporate position limits or risk controls, which explains the large drawdowns observed. Both equity curves are provided in the Figures 13 and 14. We can see that for EUR/USD it went almost instantly into the drawdown as compared to GBP/USD where we first see a really strong increase over 100,000$ before massive drawdowns consume the gains and it ends in the negative.

Although the model does identify some winning opportunities (reflected by the 35% win rates), the absence of risk constraints leads to compounding losses. This reflected raw model performance in Framework 1, underscores the importance of integrating robust position sizing and risk management strategies to harness any predictive power effectively.

RF Backtesting: Framework 2 In Framework 2, only one position can be opened at a time, as in chapter 5.2 explained. This Framework prevents the simultaneous stacking of multiple positions or opposing trades, thereby reducing risk significantly compared to Framework 1. For EUR/USD and GBP/USD only 2693 respectively 4222 trades were executed.

Description and Interpretation Table 5 summarizes the backtesting results for EUR/USD and GBP/USD under Framework 2. Notably, the backtest for EUR/USD achieves a positive total return of 56.56%, whereas GBP/USD ends with a total return of -16.63%. Both pairs maintain an overall win percentage around roughly 36%. Although this framework avoids the extreme drawdowns observed in Framework 1, the results still highlight the model's challenges in consistently identifying profitable trades.

Table 5: Random Forest Backtesting Results under Framework 2

Metric	**EUR/USD**	**GBP/USD**
Total Return (%)	56.56	-16.63
Total Return ($)	5656.43	-1662.98
Avg Return per Year (%)	3.53	-1.04
Number of Trades	2693	4222
Long Trades	1236	1821
Short Trades	1457	2401
Long Win Percentage (%)	36.65	35.97
Short Win Percentage (%)	37.27	36.15
Overall Win Percentage (%)	36.98	36.07
Date Account Reached 0	Never	Never

Similar to Framework 1, more short positions than long positions were executed. Figure 15 and Figure 16 illustrate the equity curves. For EUR/USD, the balance goes into drawdown and stays there for the next years, despite never reaching zero. Around the end of 2014 it managed to work out of drawdown starting to make profit. By contrast, GBP/USD first starts of with considerable gains reaching even over 18,000$, but then gradually loses more than it gains after this peak around 2014 and drops to the original balance at the end of 2018 / beginning of 2019, before it only remains in drawdown from 2020 and concludes with a net loss. Compared to Framework 1, the single-position constraint clearly mitigates large-scale drawdowns, though it does not guarantee overall profitability across both currency pairs. But this Framework shows a better chance of achieving a profit

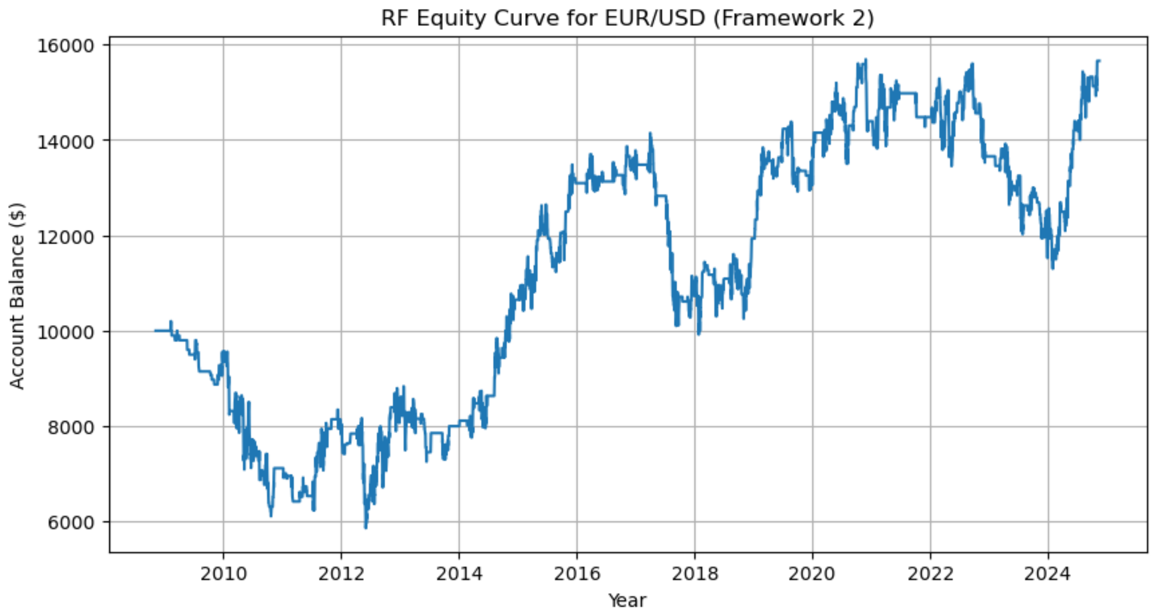

Figure 15: RF EUR/USD Equity Curve (Framework 2)

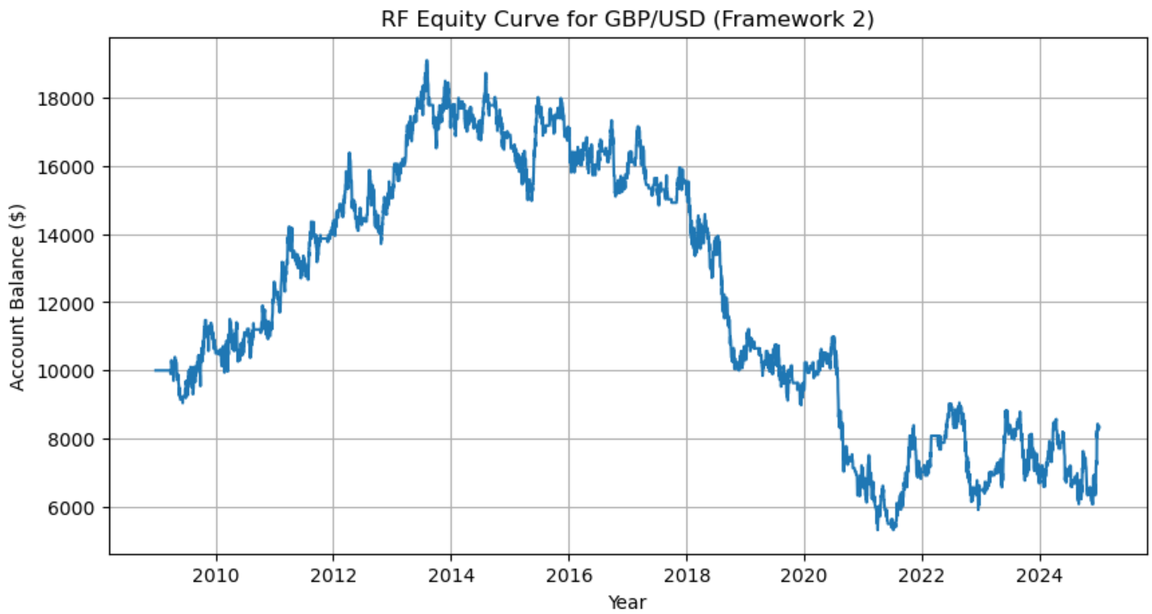

Figure 16: RF GBP/USD Equity Curve (Framework 2)

6.3 MLP Model Performance

The performance of the Multi-Layer Perceptron models for both currency pairs and all time rolling windows is presented in detail in the Tables 11 and 13 in the Appendix. As with the Random Forest results, statistical key metrics for accuracy, precision, recall, and F1-score are provided here in Table 6 for EUR/USD, and in Table 7 for GBP/USD.

Table 6: Summary Statistics for MLP Performance for EUR/USD

	Accuracy	Precision	Recall	F1-Score
min	0.154	0.107	0.239	0.096
25%	0.322	0.321	0.322	0.292
mean	0.359	0.357	0.356	0.324
median	0.356	0.355	0.354	0.328
75%	0.395	0.388	0.383	0.360
max	0.541	0.625	0.495	0.474

Table 7: Summary Statistics for MLP Performance for GBP/USD

	Accuracy	Precision	Recall	F1-Score
min	0.213	0.196	0.217	0.193
25%	0.312	0.316	0.330	0.283
mean	0.357	0.363	0.357	0.321
median	0.355	0.357	0.355	0.318
75%	0.397	0.402	0.383	0.359
max	0.534	0.645	0.562	0.531

Accuracy Accuracy values for EUR/USD range from 0.154 to 0.541, with a mean of 0.359 and a median of 0.356. GBP/USD shows a similar pattern, spanning 0.213 to 0.534, with a mean of 0.357 and a median of 0.355. These narrow differences between mean and median indicate a concentration of values near their central tendency, suggesting that extreme outliers are relatively infrequent. As shown in Figures 17 and 18, the model exhibits occasional peaks above 0.50 and dips close to 0.20, implying sensitivity to varying market regimes.

Comparison to RF: The MLP's average accuracy for both pairs is slightly lower than that of the Random Forest, although the ranges overlap substantially. Both models appear to face similar challenges when market conditions shift.

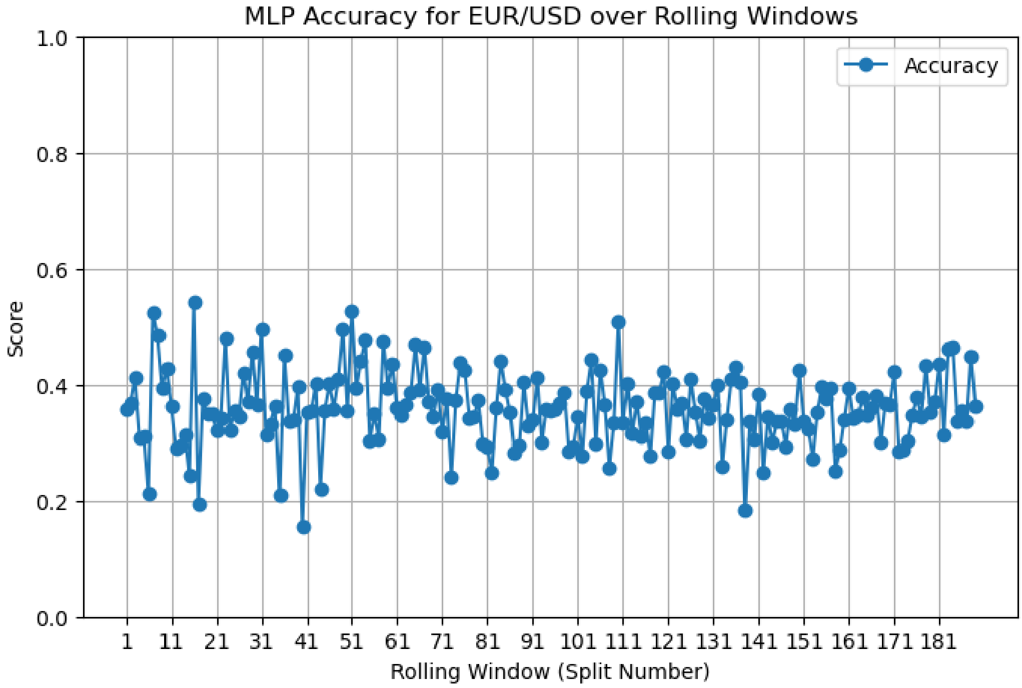

Figure 17: MLP Accuracy for EUR/USD over time rolling windows

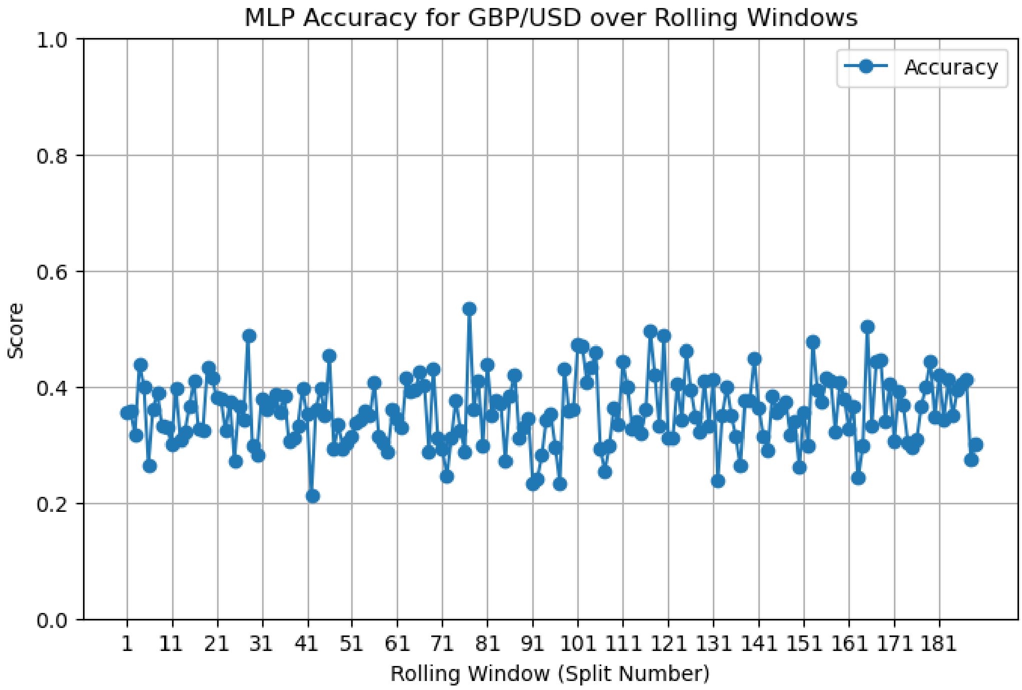

Figure 18: MLP Accuracy for GBP/USD over time rolling windows

Precision Precision for EUR/USD varies between 0.107 and 0.625, with a mean of 0.357 and a median of 0.355. For GBP/USD, it ranges from 0.196 to 0.645, with a mean of 0.363 and a median of 0.357. Figures 19 and 20 show that, while both pairs achieve occasional spikes above 0.50, the majority of values lie between 0.25 and 0.40. This indicates that although the model occasionally predicts the correct class with relatively high probability, it struggles to maintain that level across all rolling windows. This performance is moderately weak.

Comparison to RF: The MLP generally attains marginally higher mean and median precision than the RF. However, both models exhibit similar fluctuations, suggesting neither algorithm maintains consistently high precision over time.

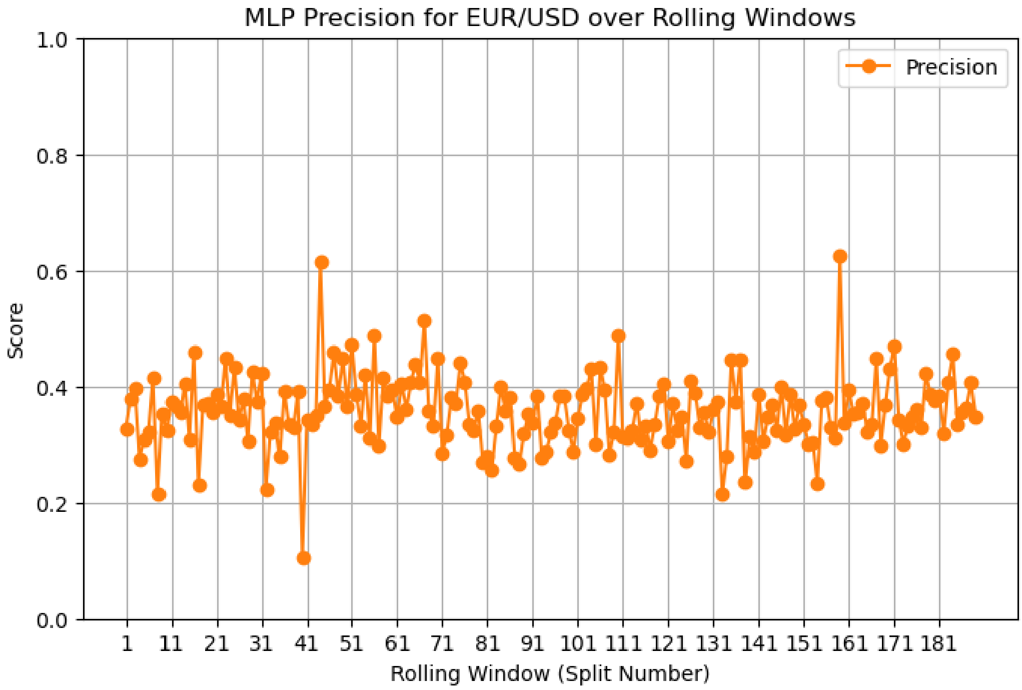

Figure 19: MLP Precision for EUR/USD over time rolling windows

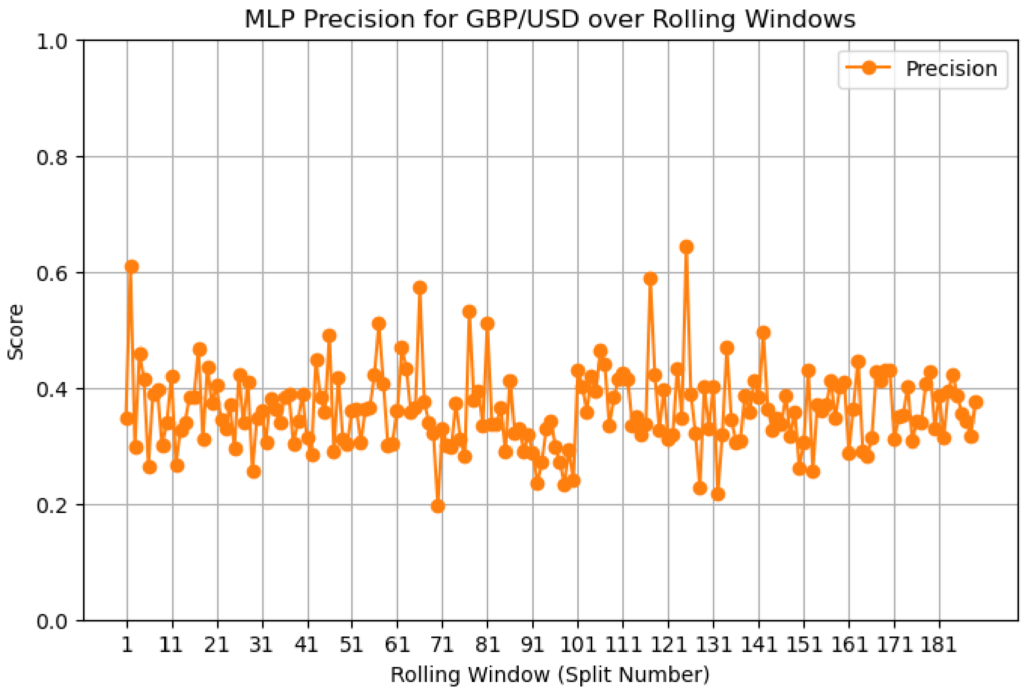

Figure 20: MLP Precision for GBP/USD over time rolling windows

Recall Recall values range from 0.239 to 0.495 for EUR/USD, with a mean of 0.356 and a median of 0.354. And from 0.217 to 0.562 for GBP/USD, with a mean of 0.357 and a median of 0.355. As depicted in Figures 21 and 22, the distribution is somewhat narrower than that observed for precision. Nevertheless, the model often fails to identify a substantial portion of the correct classes, as many windows remain below 0.40.

Comparison to RF: The MLP shows a similar overall range in recall, the mean values for both pairs compared to the random forest are not significantly higher. This indicates that the MLP identifies the true classes just as well as the RF.

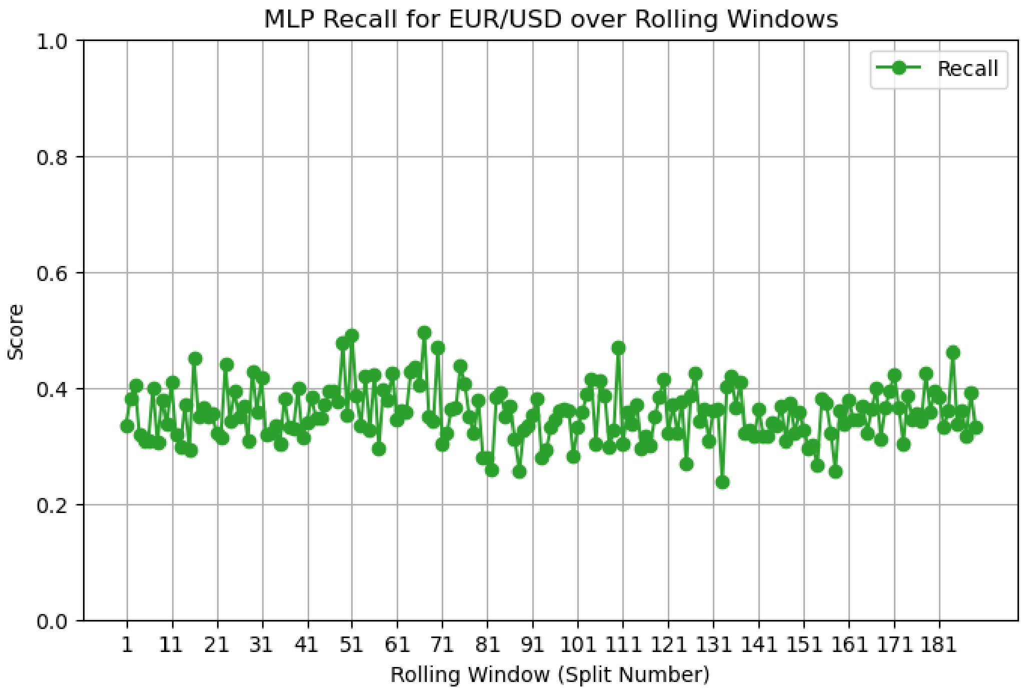

Figure 21: MLP Recall for EUR/USD over time rolling windows

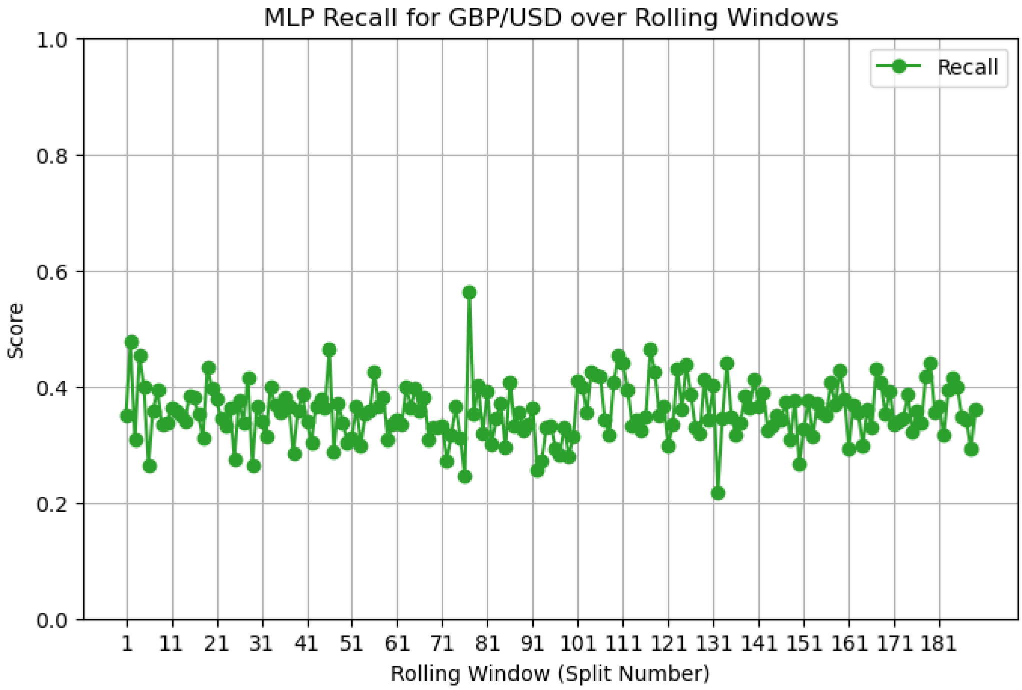

Figure 22: MLP Recall for GBP/USD over time rolling windows

F1-Score F1-scores exhibit a range of 0.096 to 0.474 for EUR/USD (mean 0.324, median 0.328) and 0.193 to 0.531 for GBP/USD (mean 0.321, median 0.318). As seen in Figures 23 and 24, the majority of F1-scores lie between roughly 0.20 and 0.40, reflecting the interplay between Precision and Recall. Peaks above 0.40 are sporadic, underlining the difficulty in balancing both metrics consistently.

Comparison to RF: The MLP's F1-scores for EUR/USD are generally in the same magnitude as the RF, although the mean and median are higher. For GBP/USD is it similar, except that the mean and median are not so much higher here compared to the RF. Both models remain in a moderate to low performance range.

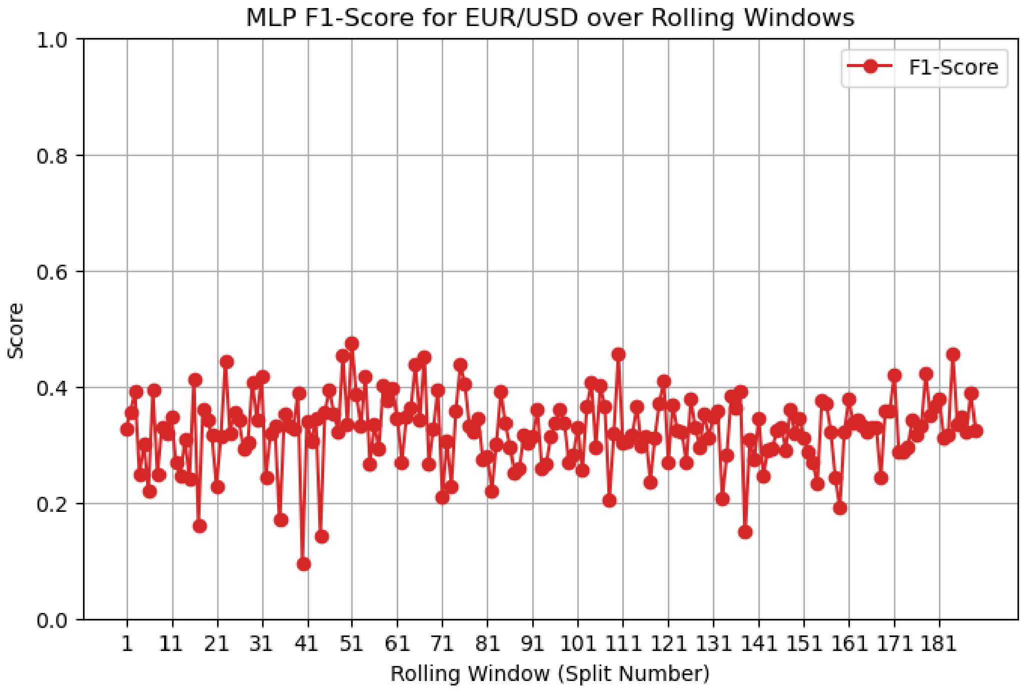

Figure 23: MLP F1-Score for EUR/USD over time rolling windows

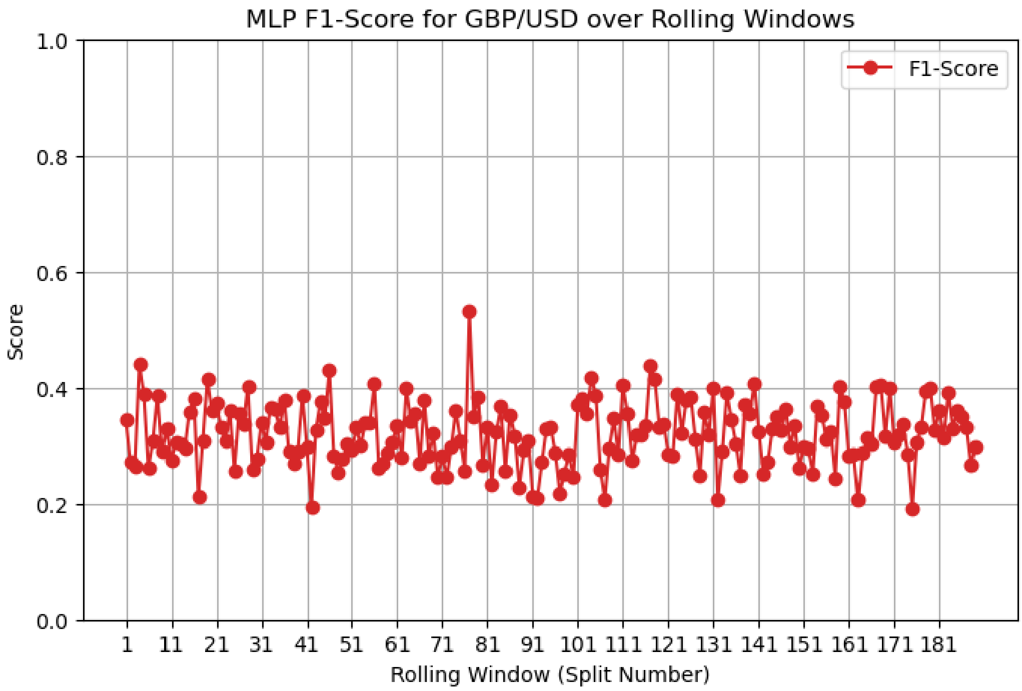

Figure 24: MLP F1-Score for GBP/USD over time rolling windows

In general, the MLP model demonstrates moderate predictive capability with considerable variability across rolling windows. While some windows achieve near 0.50 Accuracy or relatively strong Precision/Recall, others fall substantially below 0.30. This behavior suggests that, like the Random Forest, the MLP is sensitive to shifts in market regimes and changes in class distributions. Despite the MLP showing a slight edge in Recall and F1-score for certain windows, neither algorithm consistently outperforms the other in all metrics. Both appear to face similar challenges in adapting to evolving market conditions, highlighting the inherent complexity of financial time-series forecasting.

MLP Backtesting: Framework 1 In Framework 1, every signal generated by the MLP model is executed, regardless of any existing open positions, as explained in chapter 5.2.

Description and Interpretation Table 8 provides an overview of the key metrics for EUR/USD and GBP/USD under Framework 1. Notably, both currency pairs exhibit large negative total returns: -1029.59% ($-102,958.80$\$) for EUR/USD and -960.06% ($-96,005.85$\$) for GBP/USD. The account balances reached zero relatively quickly in both cases, underscoring the high risk of continuously entering positions without restrictions. Despite having slightly different distributions of long and short trades, both pairs show similar overall win rates in the mid-35% range.

Table 8: MLP Backtesting Results under Framework 1

Metric	**EUR/USD**	**GBP/USD**
Total Return (%)	-1029.59	-960.06
Total Return ($)	-102958.80	-96005.85
Avg Return per Year (%)	-64.35	-60.00
Number of Trades	58567	58044
Long Trades	29686	27471
Short Trades	28881	30573
Long Win Percentage (%)	35.17	35.50
Short Win Percentage (%)	35.42	35.09
Overall Win Percentage (%)	35.29	35.29
Date Account Reached 0	2009-06-19	2009-05-08

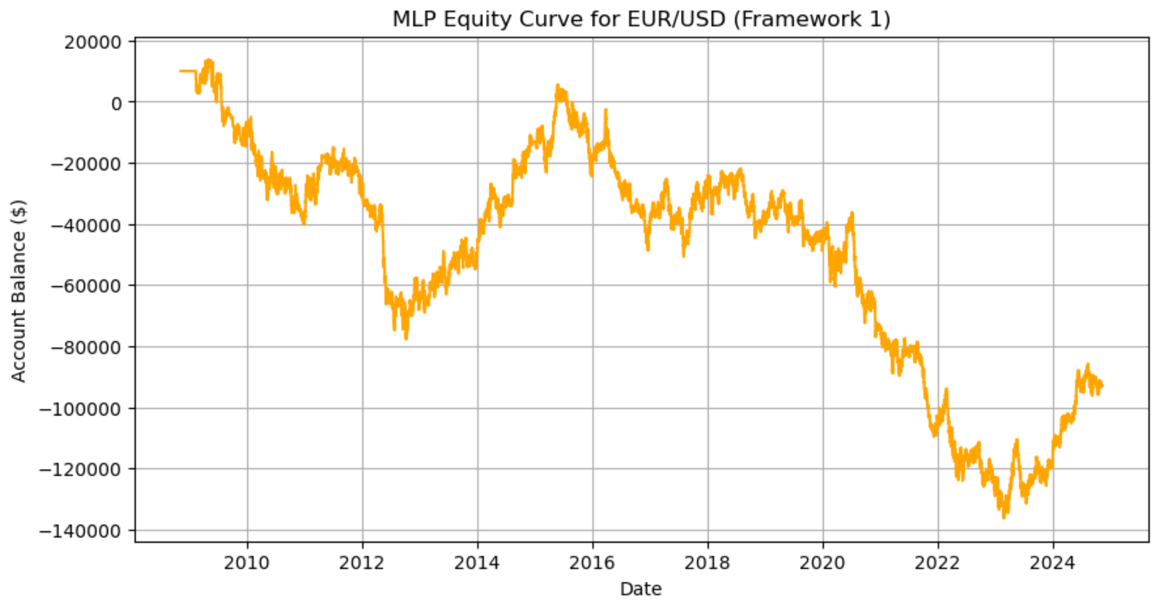

Figure 25: MLP EUR/USD Equity Curve (Framework 1)

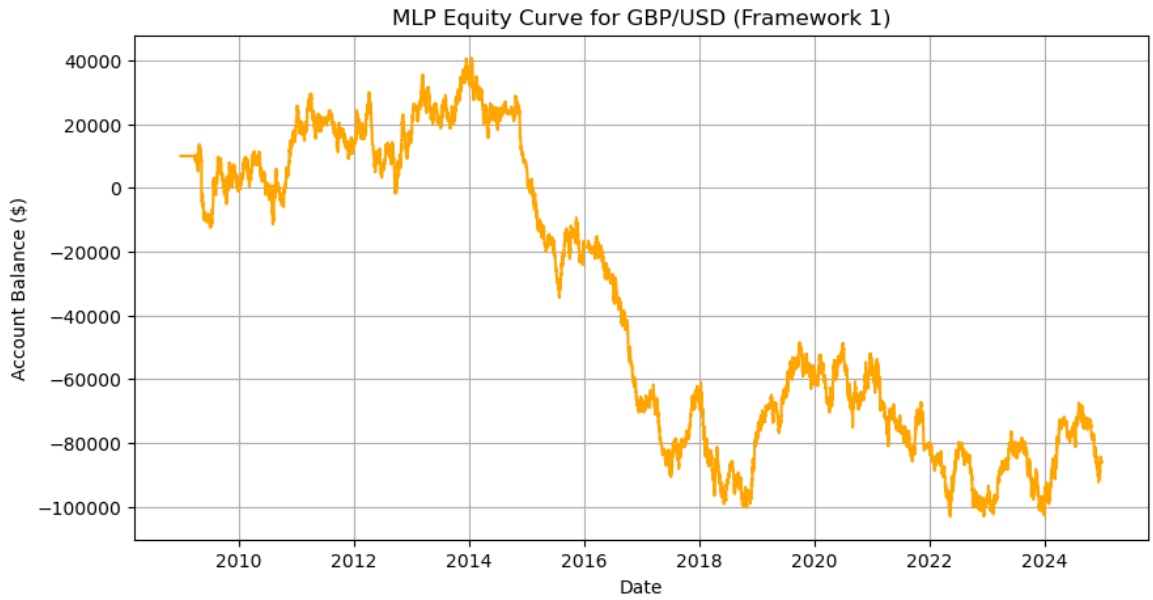

Figure 26: MLP GBP/USD Equity Curve (Framework 1)

As illustrated in Figures 25 and 26, both equity curves experience dramatic drawdowns. For EUR/USD as soon after trading begins, and it never recover to positive territory. GBP/USD starts with a drawdown, but managed to work itself out of it reaching 40,000$

at 2014, afterwards there is a sharp decline in the balance after it concludes negative. The absence of risk constraints magnifies losses when the market moves contrary to the model's predictions, particularly because multiple positions can be opened at once. While the model does show a modest overall win rate of about 35%, the consistent stacking of losing positions overwhelms any gains. This outcome highlights the need for stricter position sizing or additional filters to harness the MLP's predictive signals more effectively. Noteworthy is that here, for EUR/USD, more long positions than short positions were executed.

MLP Backtesting: Framework 2 In Framework 2, only one position at a time is permitted. Once a position is opened, the system must close it (either by stop-loss or take-profit) before initiating a new trade. This approach introduces a degree of position control and risk limitation compared to Framework 1, where multiple overlapping trades could accumulate.

Description and Interpretation Table 9 summarizes the key performance metrics for both EUR/USD and GBP/USD under Framework 2. Notably, EUR/USD achieves a total return of approximately 0.77% (+76.76$), while GBP/USD shows a more substantial total return of 32.73% (+3272.63$). Although the average annualized returns are still modest, especially for EUR/USD at around 0.05%, GBP/USD's annual average of roughly 2.05% suggests slightly more favorable conditions or model alignment for that pair.

Both pairs exhibit around 6,500 trades in total, with a nearly balanced split between long and short positions. The overall win rates, in the 36–37% range, are comparable to the Random Forest results under the same framework. However, the MLP model shows some potential for positive net growth in GBP/USD, implying that restricting positions to a single open trade at a time helps mitigate the risk of compounding losses seen in Framework 1.

Table 9: MLP Backtesting Results under Framework 2

Metric	**EUR/USD**	**GBP/USD**
Total Return (%)	0.77	32.73
Total Return ($)	76.76	3272.63
Avg Return per Year (%)	0.05	2.05
Number of Trades	6544	6546
Long Trades	3344	3155
Short Trades	3200	3391
Long Win Percentage (%)	37.05	37.21
Short Win Percentage (%)	36.13	36.42
Overall Win Percentage (%)	36.60	36.80
Date Account Reached 0	Never	Never

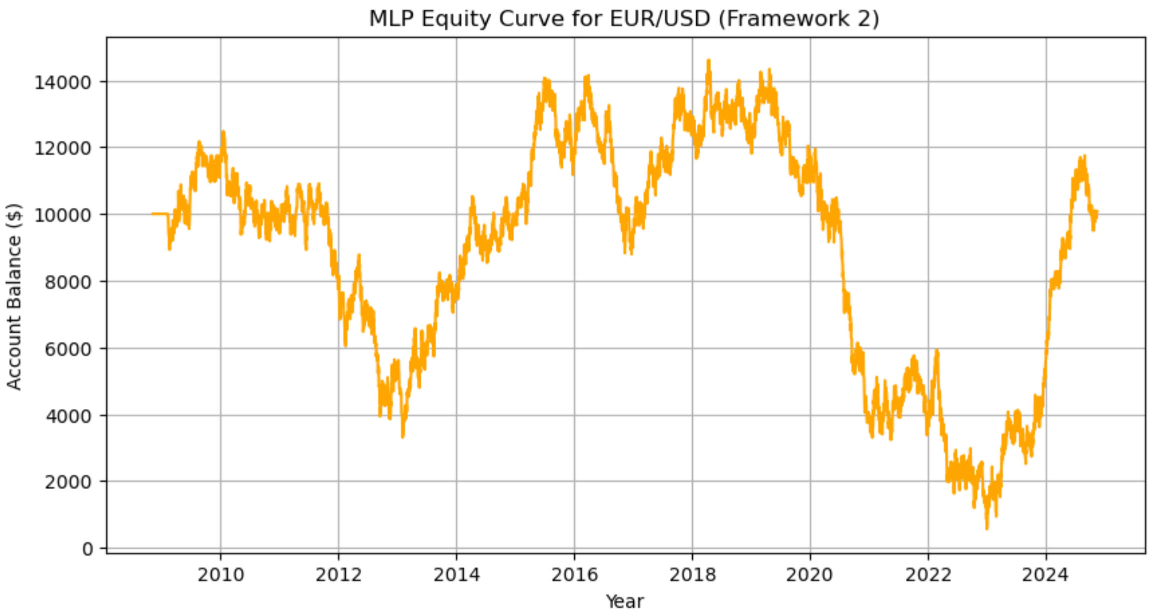

Figure 27: MLP EUR/USD Equity Curve (Framework 2)

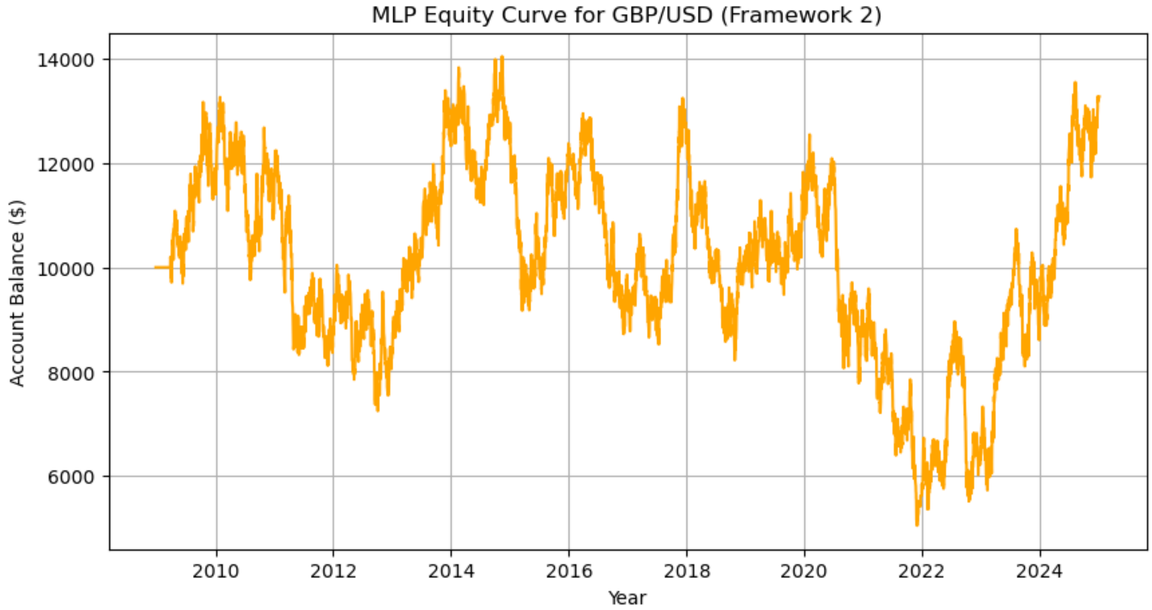

Figure 28: MLP GBP/USD Equity Curve (Framework 2)

Figure 27 shows the EUR/USD equity curve fluctuates around the initial balance and ultimately ends slightly above it. By contrast, Figure 28 illustrates a more pronounced upward trend for GBP/USD, although still subject to significant volatility. Compared to the Random Forest results under the same framework, the MLP's overall performance exhibits a similar win rate but with a slightly higher total return for GBP/USD. This finding highlights that the restriction to one open trade at a time can help the MLP model capture profitable opportunities without incurring excessive correlated losses, although the outcomes remain sensitive to market shifts and the model's inherent predictive limitations.

7 Conclusion

This study analyzed the predictive performance of machine learning and deep learning models, comparing the effectiveness of Random Forest and MLP models, for predicting trading outcomes.

Both models show only moderate performance in all metrics. In some areas even rather weak performance. Neither model turned out to be superior to the other and the choice of currency pair made no difference. Most values are between 0.20 and 0.50 which is not good for a predictive model. In addition the abrupt changes in the metrics over the different rolling windows show that the performance is not robust over the observation periods. This is despite the fact that the small time windows are intended to help the models recognize short-term changes in market dynamics. The MLP and Random Forest results suggest that both architectures can capture certain market signals but struggle when confronted with volatile or rapidly shifting market conditions.

The backtesting frameworks provide two different approaches on how to leverage these ML and DL methods in actual trading scenarios. Framework 1 reflects the raw predictive performance of the model, despite being unrealistic. Framework 2 is a more realistic approach where only one position at a time is opened, meaning that not all model predicitons are incorporated. The results have shown, that executing every signal can yield massive returns, but also provides tremendous drawdowns. Only having one position limits the risk drastically and three of four backtests using Framework 2 end up in profit.

In summary, while both RF and MLP models demonstrate moderate performance for prediction, the results underscore the importance of considering the actual circumstances where the models predictions are employed. These insights not only contribute to the growing literature at the intersection of machine learning and finance but are giving room for future improvement and for future research.

8 Appendix

Table 10: Random Forest Performance over Time Rolling Windows for EUR/ USD

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2008-11-06 to 2009-02-06	2009-02-06 to 2009-03-06	0.461	0.271	0.354	0.258	213	147	111
2008-12-06 to 2009-03-06	2009-03-06 to 2009-04-06	0.495	0.272	0.330	0.262	255	84	154
2009-01-06 to 2009-04-06	2009-04-06 to 2009-05-06	0.472	0.157	0.333	0.214	244	147	126
2009-02-06 to 2009-05-06	2009-05-06 to 2009-06-06	0.517	0.174	0.333	0.229	268	112	138
2009-03-06 to 2009-06-06	2009-06-06 to 2009-07-06	0.395	0.132	0.333	0.189	195	134	165
2009-04-06 to 2009-07-06	2009-07-06 to 2009-08-06	0.535	0.442	0.306	0.297	336	72	136
2009-05-06 to 2009-08-06	2009-08-06 to 2009-09-06	0.579	0.193	0.333	0.245	285	94	113
2009-06-06 to 2009-09-06	2009-09-06 to 2009-10-06	0.556	0.185	0.333	0.238	288	104	126
2009-07-06 to 2009-10-06	2009-10-06 to 2009-11-06	0.517	0.177	0.328	0.230	286	121	137
2009-08-06 to 2009-11-06	2009-11-06 to 2009-12-06	0.447	0.275	0.324	0.263	228	163	86
2009-09-06 to 2009-12-06	2009-12-06 to 2010-01-06	0.467	0.309	0.489	0.353	295	158	93
2009-10-06 to 2010-01-06	2010-01-06 to 2010-02-06	0.278	0.198	0.310	0.205	271	211	43
2009-11-06 to 2010-02-06	2010-02-06 to 2010-03-06	0.378	0.126	0.333	0.183	180	122	174
2009-12-06 to 2010-03-06	2010-03-06 to 2010-04-06	0.249	0.474	0.317	0.223	217	195	107
2010-01-06 to 2010-04-06	2010-04-06 to 2010-05-06	0.321	0.271	0.392	0.259	242	210	68
2010-02-06 to 2010-05-06	2010-05-06 to 2010-06-06	0.376	0.125	0.333	0.182	213	186	96
2010-03-06 to 2010-06-06	2010-06-06 to 2010-07-06	0.274	0.302	0.402	0.220	283	75	160

Continued on next page

Table 10 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2010-04-06 to 2010-07-06	2010-07-06 to 2010-08-06	0.427	0.265	0.377	0.296	235	108	200
2010-05-06 to 2010-08-06	2010-08-06 to 2010-09-06	0.457	0.533	0.337	0.267	232	145	115
2010-06-06 to 2010-09-06	2010-09-06 to 2010-10-06	0.306	0.296	0.344	0.254	274	61	185
2010-07-06 to 2010-10-06	2010-10-06 to 2010-11-06	0.328	0.242	0.335	0.178	169	177	173
2010-08-06 to 2010-11-06	2010-11-06 to 2010-12-06	0.482	0.289	0.336	0.266	247	159	92
2010-09-06 to 2010-12-06	2010-12-06 to 2011-01-06	0.467	0.156	0.333	0.212	256	152	140
2010-10-06 to 2011-01-06	2011-01-06 to 2011-02-06	0.466	0.239	0.327	0.238	242	106	152
2010-11-06 to 2011-02-06	2011-02-06 to 2011-03-06	0.357	0.210	0.283	0.241	232	120	124
2010-12-06 to 2011-03-06	2011-03-06 to 2011-04-06	0.392	0.131	0.333	0.188	212	174	155
2011-01-06 to 2011-04-06	2011-04-06 to 2011-05-06	0.423	0.306	0.363	0.249	207	151	162
2011-02-06 to 2011-05-06	2011-05-06 to 2011-06-06	0.468	0.515	0.326	0.228	239	110	145
2011-03-06 to 2011-06-06	2011-06-06 to 2011-07-06	0.456	0.152	0.333	0.209	237	170	113
2011-04-06 to 2011-07-06	2011-07-06 to 2011-08-06	0.452	0.421	0.379	0.341	228	170	122
2011-05-06 to 2011-08-06	2011-08-06 to 2011-09-06	0.514	0.445	0.404	0.404	280	143	95
2011-06-06 to 2011-09-06	2011-09-06 to 2011-10-06	0.362	0.269	0.373	0.284	250	159	111
2011-07-06 to 2011-10-06	2011-10-06 to 2011-11-06	0.464	0.487	0.339	0.221	228	122	146
2011-08-06 to 2011-11-06	2011-11-06 to 2011-12-06	0.485	0.162	0.333	0.218	253	144	125
2011-09-06 to 2011-12-06	2011-12-06 to 2012-01-06	0.504	0.221	0.315	0.258	309	146	93

Continued on next page

Table 10 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2011-10-06 to 2012-01-06	2012-01-06 to 2012-02-06	0.464	0.493	0.326	0.225	237	102	157
2011-11-06 to 2012-02-06	2012-02-06 to 2012-03-06	0.394	0.288	0.335	0.218	198	162	140
2011-12-06 to 2012-03-06	2012-03-06 to 2012-04-06	0.428	0.143	0.333	0.200	233	164	148
2012-01-06 to 2012-04-06	2012-04-06 to 2012-05-06	0.392	0.341	0.348	0.339	217	129	126
2012-02-06 to 2012-05-06	2012-05-06 to 2012-06-06	0.152	0.168	0.305	0.097	237	215	86
2012-03-06 to 2012-06-06	2012-06-06 to 2012-07-06	0.390	0.212	0.304	0.223	225	175	116
2012-04-06 to 2012-07-06	2012-07-06 to 2012-08-06	0.325	0.284	0.354	0.260	217	151	125
2012-05-06 to 2012-08-06	2012-08-06 to 2012-09-06	0.484	0.308	0.390	0.342	253	169	119
2012-06-06 to 2012-09-06	2012-09-06 to 2012-10-06	0.262	0.289	0.376	0.193	227	100	166
2012-07-06 to 2012-10-06	2012-10-06 to 2012-11-06	0.426	0.296	0.386	0.317	190	190	137
2012-08-06 to 2012-11-06	2012-11-06 to 2012-12-06	0.367	0.388	0.352	0.301	199	143	179
2012-09-06 to 2012-12-06	2012-12-06 to 2013-01-06	0.335	0.571	0.410	0.322	188	104	188
2012-10-06 to 2013-01-06	2013-01-06 to 2013-02-06	0.240	0.192	0.339	0.182	285	101	160
2012-11-06 to 2013-02-06	2013-02-06 to 2013-03-06	0.544	0.242	0.339	0.278	274	129	73
2012-12-06 to 2013-03-06	2013-03-06 to 2013-04-06	0.435	0.315	0.424	0.341	214	151	155
2013-01-06 to 2013-04-06	2013-04-06 to 2013-05-06	0.583	0.194	0.333	0.246	288	80	126
2013-02-06 to 2013-05-06	2013-05-06 to 2013-06-06	0.416	0.335	0.339	0.234	228	146	169
2013-03-06 to 2013-06-06	2013-06-06 to 2013-07-06	0.516	0.244	0.319	0.243	270	162	62

Continued on next page

Table 10 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2013-04-06 to 2013-07-06	2013-07-06 to 2013-08-06	0.537	0.488	0.361	0.281	263	69	186
2013-05-06 to 2013-08-06	2013-08-06 to 2013-09-06	0.349	0.116	0.333	0.173	190	211	143
2013-06-06 to 2013-09-06	2013-09-06 to 2013-10-06	0.356	0.262	0.390	0.280	231	88	153
2013-07-06 to 2013-10-06	2013-10-06 to 2013-11-06	0.467	0.254	0.336	0.251	259	147	136
2013-08-06 to 2013-11-06	2013-11-06 to 2013-12-06	0.513	0.171	0.333	0.226	269	107	148
2013-09-06 to 2013-12-06	2013-12-06 to 2014-01-06	0.460	0.361	0.344	0.237	210	108	143
2013-10-06 to 2014-01-06	2014-01-06 to 2014-02-06	0.527	0.176	0.333	0.230	289	158	101
2013-11-06 to 2014-02-06	2014-02-06 to 2014-03-06	0.395	0.132	0.332	0.189	189	107	180
2013-12-06 to 2014-03-06	2014-03-06 to 2014-04-06	0.394	0.328	0.377	0.278	257	151	89
2014-01-06 to 2014-04-06	2014-04-06 to 2014-05-06	0.500	0.167	0.333	0.222	259	92	167
2014-02-06 to 2014-05-06	2014-05-06 to 2014-06-06	0.276	0.305	0.359	0.236	265	173	106
2014-03-06 to 2014-06-06	2014-06-06 to 2014-07-06	0.451	0.295	0.387	0.326	217	142	113
2014-04-06 to 2014-07-06	2014-07-06 to 2014-08-06	0.426	0.142	0.333	0.199	231	240	71
2014-05-06 to 2014-08-06	2014-08-06 to 2014-09-06	0.477	0.508	0.487	0.470	228	187	105
2014-06-06 to 2014-09-06	2014-09-06 to 2014-10-06	0.369	0.456	0.338	0.189	180	178	135
2014-07-06 to 2014-10-06	2014-10-06 to 2014-11-06	0.306	0.305	0.304	0.300	231	155	156
2014-08-06 to 2014-11-06	2014-11-06 to 2014-12-06	0.322	0.389	0.398	0.311	253	147	100
2014-09-06 to 2014-12-06	2014-12-06 to 2015-01-06	0.329	0.137	0.312	0.175	170	171	143

Continued on next page

Table 10 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2014-10-06 to 2015-01-06	2015-01-06 to 2015-02-06	0.354	0.238	0.302	0.255	235	205	108
2014-11-06 to 2015-02-06	2015-02-06 to 2015-03-06	0.492	0.460	0.415	0.385	231	176	69
2014-12-06 to 2015-03-06	2015-03-06 to 2015-04-06	0.347	0.221	0.366	0.275	169	133	193
2015-01-06 to 2015-04-06	2015-04-06 to 2015-05-06	0.289	0.304	0.274	0.235	205	148	166
2015-02-06 to 2015-05-06	2015-05-06 to 2015-06-06	0.392	0.302	0.341	0.297	221	145	152
2015-03-06 to 2015-06-06	2015-06-06 to 2015-07-06	0.387	0.312	0.300	0.299	250	143	100
2015-04-06 to 2015-07-06	2015-07-06 to 2015-08-06	0.427	0.169	0.316	0.211	245	151	145
2015-05-06 to 2015-08-06	2015-08-06 to 2015-09-06	0.512	0.373	0.444	0.382	220	127	147
2015-06-06 to 2015-09-06	2015-09-06 to 2015-10-06	0.449	0.320	0.408	0.340	218	131	168
2015-07-06 to 2015-10-06	2015-10-06 to 2015-11-06	0.363	0.240	0.326	0.276	221	174	148
2015-08-06 to 2015-11-06	2015-11-06 to 2015-12-06	0.462	0.352	0.411	0.339	247	140	85
2015-09-06 to 2015-12-06	2015-12-06 to 2016-01-06	0.448	0.150	0.333	0.207	227	119	161
2015-10-06 to 2016-01-06	2016-01-06 to 2016-02-06	0.457	0.152	0.333	0.209	239	84	200
2015-11-06 to 2016-02-06	2016-02-06 to 2016-03-06	0.484	0.418	0.435	0.366	221	186	68
2015-12-06 to 2016-03-06	2016-03-06 to 2016-04-06	0.492	0.307	0.394	0.342	296	131	116
2016-01-06 to 2016-04-06	2016-04-06 to 2016-05-06	0.362	0.214	0.365	0.268	202	96	219
2016-02-06 to 2016-05-06	2016-05-06 to 2016-06-06	0.482	0.246	0.348	0.262	237	138	117
2016-03-06 to 2016-06-06	2016-06-06 to 2016-07-06	0.481	0.494	0.335	0.223	249	100	169

Continued on next page

Table 10 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2016-04-06 to 2016-07-06	2016-07-06 to 2016-08-06	0.470	0.157	0.333	0.213	244	117	158
2016-05-06 to 2016-08-06	2016-08-06 to 2016-09-06	0.429	0.464	0.347	0.291	246	99	168
2016-06-06 to 2016-09-06	2016-09-06 to 2016-10-06	0.395	0.132	0.333	0.189	204	168	145
2016-07-06 to 2016-10-06	2016-10-06 to 2016-11-06	0.457	0.345	0.381	0.285	205	139	151
2016-08-06 to 2016-11-06	2016-11-06 to 2016-12-06	0.549	0.238	0.331	0.249	292	134	95
2016-09-06 to 2016-12-06	2016-12-06 to 2017-01-06	0.380	0.271	0.341	0.277	264	107	177
2016-10-06 to 2017-01-06	2017-01-06 to 2017-02-06	0.442	0.147	0.333	0.204	219	156	121
2016-11-06 to 2017-02-06	2017-02-06 to 2017-03-06	0.402	0.134	0.333	0.191	191	166	118
2016-12-06 to 2017-03-06	2017-03-06 to 2017-04-06	0.481	0.305	0.383	0.337	258	171	114
2017-01-06 to 2017-04-06	2017-04-06 to 2017-05-06	0.367	0.245	0.382	0.280	242	82	167
2017-02-06 to 2017-05-06	2017-05-06 to 2017-06-06	0.413	0.138	0.333	0.195	214	123	181
2017-03-06 to 2017-06-06	2017-06-06 to 2017-07-06	0.416	0.139	0.333	0.196	216	148	155
2017-04-06 to 2017-07-06	2017-07-06 to 2017-08-06	0.194	0.282	0.341	0.129	223	88	183
2017-05-06 to 2017-08-06	2017-08-06 to 2017-09-06	0.275	0.287	0.295	0.275	257	158	126
2017-06-06 to 2017-09-06	2017-09-06 to 2017-10-06	0.377	0.251	0.334	0.286	218	172	127
2017-07-06 to 2017-10-06	2017-10-06 to 2017-11-06	0.436	0.146	0.329	0.202	217	146	128
2017-08-06 to 2017-11-06	2017-11-06 to 2017-12-06	0.401	0.182	0.299	0.221	242	111	168
2017-09-06 to 2017-12-06	2017-12-06 to 2018-01-06	0.307	0.374	0.326	0.268	182	100	206

Continued on next page

Table 10 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2017-10-06 to 2018-01-06	2018-01-06 to 2018-02-06	0.283	0.187	0.285	0.225	203	139	178
2017-11-06 to 2018-02-06	2018-02-06 to 2018-03-06	0.432	0.337	0.325	0.305	241	117	117
2017-12-06 to 2018-03-06	2018-03-06 to 2018-04-06	0.535	0.179	0.332	0.233	290	133	117
2018-01-06 to 2018-04-06	2018-04-06 to 2018-05-06	0.362	0.256	0.289	0.222	206	180	86
2018-02-06 to 2018-05-06	2018-05-06 to 2018-06-06	0.385	0.242	0.335	0.257	211	189	140
2018-03-06 to 2018-06-06	2018-06-06 to 2018-07-06	0.355	0.217	0.304	0.251	226	148	145
2018-04-06 to 2018-07-06	2018-07-06 to 2018-08-06	0.421	0.367	0.355	0.246	202	170	122
2018-05-06 to 2018-08-06	2018-08-06 to 2018-09-06	0.461	0.431	0.448	0.385	238	160	144
2018-06-06 to 2018-09-06	2018-09-06 to 2018-10-06	0.329	0.263	0.308	0.268	197	133	163
2018-07-06 to 2018-10-06	2018-10-06 to 2018-11-06	0.449	0.271	0.363	0.307	245	135	137
2018-08-06 to 2018-11-06	2018-11-06 to 2018-12-06	0.359	0.239	0.360	0.281	209	158	154
2018-09-06 to 2018-12-06	2018-12-06 to 2019-01-06	0.469	0.458	0.401	0.314	182	107	174
2018-10-06 to 2019-01-06	2019-01-06 to 2019-02-06	0.515	0.623	0.397	0.360	265	187	94
2018-11-06 to 2019-02-06	2019-02-06 to 2019-03-06	0.303	0.365	0.358	0.281	243	156	76
2018-12-06 to 2019-03-06	2019-03-06 to 2019-04-06	0.477	0.318	0.336	0.272	258	117	143
2019-01-06 to 2019-04-06	2019-04-06 to 2019-05-06	0.424	0.263	0.359	0.303	244	129	120
2019-02-06 to 2019-05-06	2019-05-06 to 2019-06-06	0.475	0.353	0.352	0.258	254	162	127
2019-03-06 to 2019-06-06	2019-06-06 to 2019-07-06	0.483	0.295	0.363	0.318	258	123	114

Continued on next page

Table 10 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2019-04-06 to 2019-07-06	2019-07-06 to 2019-08-06	0.448	0.286	0.399	0.331	233	157	128
2019-05-06 to 2019-08-06	2019-08-06 to 2019-09-06	0.302	0.231	0.376	0.240	272	178	90
2019-06-06 to 2019-09-06	2019-09-06 to 2019-10-06	0.373	0.234	0.345	0.278	213	149	105
2019-07-06 to 2019-10-06	2019-10-06 to 2019-11-06	0.463	0.155	0.332	0.211	252	127	163
2019-08-06 to 2019-11-06	2019-11-06 to 2019-12-06	0.521	0.174	0.333	0.229	271	132	117
2019-09-06 to 2019-12-06	2019-12-06 to 2020-01-06	0.443	0.290	0.415	0.341	181	138	139
2019-10-06 to 2020-01-06	2020-01-06 to 2020-02-06	0.488	0.367	0.371	0.300	247	197	101
2019-11-06 to 2020-02-06	2020-02-06 to 2020-03-06	0.353	0.137	0.325	0.182	180	133	183
2019-12-06 to 2020-03-06	2020-03-06 to 2020-04-06	0.375	0.362	0.388	0.344	219	176	98
2020-01-06 to 2020-04-06	2020-04-06 to 2020-05-06	0.319	0.366	0.327	0.303	263	138	119
2020-02-06 to 2020-05-06	2020-05-06 to 2020-06-06	0.338	0.327	0.368	0.287	227	63	227
2020-03-06 to 2020-06-06	2020-06-06 to 2020-07-06	0.371	0.156	0.328	0.185	185	162	144
2020-04-06 to 2020-07-06	2020-07-06 to 2020-08-06	0.235	0.202	0.341	0.199	240	84	216
2020-05-06 to 2020-08-06	2020-08-06 to 2020-09-06	0.417	0.467	0.364	0.247	192	150	152
2020-06-06 to 2020-09-06	2020-09-06 to 2020-10-06	0.488	0.410	0.384	0.308	234	144	140
2020-07-06 to 2020-10-06	2020-10-06 to 2020-11-06	0.411	0.276	0.357	0.277	213	143	187
2020-08-06 to 2020-11-06	2020-11-06 to 2020-12-06	0.339	0.193	0.317	0.238	207	74	194
2020-09-06 to 2020-12-06	2020-12-06 to 2021-01-06	0.438	0.393	0.348	0.238	214	106	187

Continued on next page

Table 10 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2020-10-06 to 2021-01-06	2021-01-06 to 2021-02-06	0.302	0.176	0.301	0.220	194	212	117
2020-11-06 to 2021-02-06	2021-02-06 to 2021-03-06	0.374	0.358	0.371	0.291	165	169	142
2020-12-06 to 2021-03-06	2021-03-06 to 2021-04-06	0.293	0.239	0.340	0.196	222	155	141
2021-01-06 to 2021-04-06	2021-04-06 to 2021-05-06	0.466	0.302	0.363	0.266	232	111	174
2021-02-06 to 2021-05-06	2021-05-06 to 2021-06-06	0.327	0.225	0.369	0.249	195	133	168
2021-03-06 to 2021-06-06	2021-06-06 to 2021-07-06	0.463	0.189	0.329	0.232	245	159	110
2021-04-06 to 2021-07-06	2021-07-06 to 2021-08-06	0.482	0.161	0.333	0.217	261	170	111
2021-05-06 to 2021-08-06	2021-08-06 to 2021-09-06	0.544	0.181	0.333	0.235	267	79	145
2021-06-06 to 2021-09-06	2021-09-06 to 2021-10-06	0.439	0.219	0.358	0.263	230	200	87
2021-07-06 to 2021-10-06	2021-10-06 to 2021-11-06	0.538	0.179	0.333	0.233	278	119	120
2021-08-06 to 2021-11-06	2021-11-06 to 2021-12-06	0.481	0.298	0.335	0.247	242	161	94
2021-09-06 to 2021-12-06	2021-12-06 to 2022-01-06	0.396	0.132	0.333	0.189	215	185	143
2021-10-06 to 2022-01-06	2022-01-06 to 2022-02-06	0.394	0.426	0.372	0.361	194	165	141
2021-11-06 to 2022-02-06	2022-02-06 to 2022-03-06	0.357	0.240	0.264	0.238	217	198	58
2021-12-06 to 2022-03-06	2022-03-06 to 2022-04-06	0.281	0.094	0.333	0.146	222	152	166
2022-01-06 to 2022-04-06	2022-04-06 to 2022-05-06	0.253	0.084	0.333	0.135	222	165	131
2022-02-06 to 2022-05-06	2022-05-06 to 2022-06-06	0.446	0.435	0.419	0.418	206	139	148
2022-03-06 to 2022-06-06	2022-06-06 to 2022-07-06	0.391	0.422	0.329	0.218	207	209	101

Continued on next page

Table 10 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2022-04-06 to 2022-07-06	2022-07-06 to 2022-08-06	0.432	0.274	0.356	0.291	223	165	131
2022-05-06 to 2022-08-06	2022-08-06 to 2022-09-06	0.335	0.316	0.345	0.299	198	164	145
2022-06-06 to 2022-09-06	2022-09-06 to 2022-10-06	0.366	0.257	0.316	0.262	224	157	127
2022-07-06 to 2022-10-06	2022-10-06 to 2022-11-06	0.423	0.473	0.339	0.208	203	122	160
2022-08-06 to 2022-11-06	2022-11-06 to 2022-12-06	0.220	0.133	0.275	0.143	237	133	152
2022-09-06 to 2022-12-06	2022-12-06 to 2023-01-06	0.536	0.179	0.333	0.233	294	93	161
2022-10-06 to 2023-01-06	2023-01-06 to 2023-02-06	0.491	0.170	0.331	0.225	247	136	116
2022-11-06 to 2023-02-06	2023-02-06 to 2023-03-06	0.403	0.467	0.337	0.199	189	163	122
2022-12-06 to 2023-03-06	2023-03-06 to 2023-04-06	0.365	0.403	0.349	0.304	205	113	225
2023-01-06 to 2023-04-06	2023-04-06 to 2023-05-06	0.323	0.154	0.339	0.176	175	157	164
2023-02-06 to 2023-05-06	2023-05-06 to 2023-06-06	0.422	0.523	0.395	0.332	223	222	72
2023-03-06 to 2023-06-06	2023-06-06 to 2023-07-06	0.410	0.321	0.324	0.208	221	121	178
2023-04-06 to 2023-07-06	2023-07-06 to 2023-08-06	0.342	0.321	0.351	0.277	235	104	155
2023-05-06 to 2023-08-06	2023-08-06 to 2023-09-06	0.431	0.144	0.332	0.201	234	183	123
2023-06-06 to 2023-09-06	2023-09-06 to 2023-10-06	0.405	0.274	0.358	0.299	210	143	166
2023-07-06 to 2023-10-06	2023-10-06 to 2023-11-06	0.405	0.171	0.318	0.200	210	111	173
2023-08-06 to 2023-11-06	2023-11-06 to 2023-12-06	0.410	0.186	0.318	0.209	226	163	133
2023-09-06 to 2023-12-06	2023-12-06 to 2024-01-06	0.324	0.546	0.349	0.261	167	142	176

Continued on next page

Table 10 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2023-10-06 to 2024-01-06	2024-01-06 to 2024-02-06	0.421	0.303	0.349	0.302	241	174	105
2023-11-06 to 2024-02-06	2024-02-06 to 2024-03-06	0.573	0.358	0.404	0.373	272	72	153
2023-12-06 to 2024-03-06	2024-03-06 to 2024-04-06	0.357	0.401	0.332	0.237	191	191	139
2024-01-06 to 2024-04-06	2024-04-06 to 2024-05-06	0.389	0.279	0.387	0.297	231	126	136
2024-02-06 to 2024-05-06	2024-05-06 to 2024-06-06	0.481	0.433	0.455	0.380	247	143	153
2024-03-06 to 2024-06-06	2024-06-06 to 2024-07-06	0.338	0.429	0.328	0.217	170	165	159
2024-04-06 to 2024-07-06	2024-07-06 to 2024-08-06	0.431	0.441	0.429	0.373	233	133	152
2024-05-06 to 2024-08-06	2024-08-06 to 2024-09-06	0.399	0.259	0.374	0.298	266	116	162
2024-06-06 to 2024-09-06	2024-09-06 to 2024-10-06	0.481	0.300	0.393	0.335	220	122	130
2024-07-06 to 2024-10-06	2024-10-06 to 2024-11-06	0.446	0.280	0.394	0.325	245	167	128

Table 11: MLP Performance over Rolling Windows for EUR/ USD

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2008-11-06 to 2009-02-06	2009-02-06 to 2009-03-06	0.357	0.328	0.335	0.326	213	147	111
2008-12-06 to 2009-03-06	2009-03-06 to 2009-04-06	0.367	0.379	0.382	0.356	255	84	154
2009-01-06 to 2009-04-06	2009-04-06 to 2009-05-06	0.412	0.397	0.404	0.391	244	147	126
2009-02-06 to 2009-05-06	2009-05-06 to 2009-06-06	0.309	0.276	0.318	0.248	268	112	138
2009-03-06 to 2009-06-06	2009-06-06 to 2009-07-06	0.310	0.307	0.308	0.301	195	134	165
2009-04-06 to 2009-07-06	2009-07-06 to 2009-08-06	0.211	0.321	0.309	0.220	336	72	136
2009-05-06 to 2009-08-06	2009-08-06 to 2009-09-06	0.524	0.414	0.399	0.393	285	94	113
2009-06-06 to 2009-09-06	2009-09-06 to 2009-10-06	0.485	0.215	0.305	0.249	288	104	126
2009-07-06 to 2009-10-06	2009-10-06 to 2009-11-06	0.395	0.353	0.380	0.330	286	121	137
2009-08-06 to 2009-11-06	2009-11-06 to 2009-12-06	0.428	0.324	0.336	0.320	228	163	86
2009-09-06 to 2009-12-06	2009-12-06 to 2010-01-06	0.363	0.373	0.409	0.348	295	158	93
2009-10-06 to 2010-01-06	2010-01-06 to 2010-02-06	0.291	0.366	0.318	0.269	271	211	43
2009-11-06 to 2010-02-06	2010-02-06 to 2010-03-06	0.294	0.355	0.298	0.245	180	122	174
2009-12-06 to 2010-03-06	2010-03-06 to 2010-04-06	0.314	0.405	0.370	0.309	217	195	107
2010-01-06 to 2010-04-06	2010-04-06 to 2010-05-06	0.242	0.309	0.294	0.241	242	210	68
2010-02-06 to 2010-05-06	2010-05-06 to 2010-06-06	0.541	0.459	0.452	0.413	213	186	96
2010-03-06 to 2010-06-06	2010-06-06 to 2010-07-06	0.193	0.229	0.351	0.161	283	75	160

Continued on next page

Table 11 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2010-04-06 to 2010-07-06	2010-07-06 to 2010-08-06	0.376	0.367	0.366	0.361	235	108	200
2010-05-06 to 2010-08-06	2010-08-06 to 2010-09-06	0.350	0.371	0.349	0.341	232	145	115
2010-06-06 to 2010-09-06	2010-09-06 to 2010-10-06	0.350	0.355	0.355	0.316	274	61	185
2010-07-06 to 2010-10-06	2010-10-06 to 2010-11-06	0.322	0.385	0.322	0.228	169	177	173
2010-08-06 to 2010-11-06	2010-11-06 to 2010-12-06	0.343	0.365	0.314	0.314	247	159	92
2010-09-06 to 2010-12-06	2010-12-06 to 2011-01-06	0.480	0.448	0.441	0.444	256	152	140
2010-10-06 to 2011-01-06	2011-01-06 to 2011-02-06	0.322	0.351	0.342	0.320	242	106	152
2010-11-06 to 2011-02-06	2011-02-06 to 2011-03-06	0.355	0.434	0.393	0.356	232	120	124
2010-12-06 to 2011-03-06	2011-03-06 to 2011-04-06	0.344	0.341	0.350	0.341	212	174	155
2011-01-06 to 2011-04-06	2011-04-06 to 2011-05-06	0.421	0.378	0.367	0.292	207	151	162
2011-02-06 to 2011-05-06	2011-05-06 to 2011-06-06	0.370	0.307	0.308	0.303	239	110	145
2011-03-06 to 2011-06-06	2011-06-06 to 2011-07-06	0.458	0.426	0.428	0.407	237	170	113
2011-04-06 to 2011-07-06	2011-07-06 to 2011-08-06	0.365	0.373	0.359	0.342	228	170	122
2011-05-06 to 2011-08-06	2011-08-06 to 2011-09-06	0.496	0.421	0.416	0.418	280	143	95
2011-06-06 to 2011-09-06	2011-09-06 to 2011-10-06	0.313	0.223	0.319	0.244	250	159	111
2011-07-06 to 2011-10-06	2011-10-06 to 2011-11-06	0.333	0.321	0.320	0.319	228	122	146
2011-08-06 to 2011-11-06	2011-11-06 to 2011-12-06	0.364	0.336	0.333	0.333	253	144	125
2011-09-06 to 2011-12-06	2011-12-06 to 2012-01-06	0.210	0.280	0.303	0.170	309	146	93

Continued on next page

Table 11 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2011-10-06 to 2012-01-06	2012-01-06 to 2012-02-06	0.452	0.391	0.381	0.353	237	102	157
2011-11-06 to 2012-02-06	2012-02-06 to 2012-03-06	0.336	0.334	0.332	0.332	198	162	140
2011-12-06 to 2012-03-06	2012-03-06 to 2012-04-06	0.339	0.328	0.329	0.328	233	164	148
2012-01-06 to 2012-04-06	2012-04-06 to 2012-05-06	0.396	0.391	0.398	0.388	217	129	126
2012-02-06 to 2012-05-06	2012-05-06 to 2012-06-06	0.154	0.107	0.314	0.096	237	215	86
2012-03-06 to 2012-06-06	2012-06-06 to 2012-07-06	0.353	0.343	0.338	0.339	225	175	116
2012-04-06 to 2012-07-06	2012-07-06 to 2012-08-06	0.355	0.333	0.383	0.305	217	151	125
2012-05-06 to 2012-08-06	2012-08-06 to 2012-09-06	0.403	0.349	0.348	0.344	253	169	119
2012-06-06 to 2012-09-06	2012-09-06 to 2012-10-06	0.221	0.615	0.348	0.142	227	100	166
2012-07-06 to 2012-10-06	2012-10-06 to 2012-11-06	0.356	0.366	0.371	0.356	190	190	137
2012-08-06 to 2012-11-06	2012-11-06 to 2012-12-06	0.401	0.395	0.395	0.395	199	143	179
2012-09-06 to 2012-12-06	2012-12-06 to 2013-01-06	0.358	0.459	0.394	0.352	188	104	188
2012-10-06 to 2013-01-06	2013-01-06 to 2013-02-06	0.410	0.384	0.377	0.322	285	101	160
2012-11-06 to 2013-02-06	2013-02-06 to 2013-03-06	0.496	0.449	0.476	0.454	274	129	73
2012-12-06 to 2013-03-06	2013-03-06 to 2013-04-06	0.356	0.365	0.352	0.335	214	151	155
2013-01-06 to 2013-04-06	2013-04-06 to 2013-05-06	0.526	0.472	0.491	0.474	288	80	126
2013-02-06 to 2013-05-06	2013-05-06 to 2013-06-06	0.394	0.386	0.385	0.385	228	146	169
2013-03-06 to 2013-06-06	2013-06-06 to 2013-07-06	0.441	0.332	0.333	0.332	270	162	62

Continued on next page

Table 11 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2013-04-06 to 2013-07-06	2013-07-06 to 2013-08-06	0.477	0.421	0.419	0.418	263	69	186
2013-05-06 to 2013-08-06	2013-08-06 to 2013-09-06	0.303	0.312	0.327	0.267	190	211	143
2013-06-06 to 2013-09-06	2013-09-06 to 2013-10-06	0.350	0.488	0.423	0.334	231	88	153
2013-07-06 to 2013-10-06	2013-10-06 to 2013-11-06	0.306	0.297	0.295	0.292	259	147	136
2013-08-06 to 2013-11-06	2013-11-06 to 2013-12-06	0.475	0.414	0.398	0.402	269	107	148
2013-09-06 to 2013-12-06	2013-12-06 to 2014-01-06	0.395	0.383	0.378	0.375	210	108	143
2013-10-06 to 2014-01-06	2014-01-06 to 2014-02-06	0.436	0.393	0.426	0.396	289	158	101
2013-11-06 to 2014-02-06	2014-02-06 to 2014-03-06	0.359	0.348	0.344	0.345	189	107	180
2013-12-06 to 2014-03-06	2014-03-06 to 2014-04-06	0.348	0.404	0.360	0.270	257	151	89
2014-01-06 to 2014-04-06	2014-04-06 to 2014-05-06	0.365	0.360	0.359	0.346	259	92	167
2014-02-06 to 2014-05-06	2014-05-06 to 2014-06-06	0.386	0.406	0.429	0.362	265	173	106
2014-03-06 to 2014-06-06	2014-06-06 to 2014-07-06	0.468	0.438	0.437	0.437	217	142	113
2014-04-06 to 2014-07-06	2014-07-06 to 2014-08-06	0.391	0.408	0.405	0.343	231	240	71
2014-05-06 to 2014-08-06	2014-08-06 to 2014-09-06	0.465	0.513	0.495	0.452	228	187	105
2014-06-06 to 2014-09-06	2014-09-06 to 2014-10-06	0.371	0.357	0.350	0.268	180	178	135
2014-07-06 to 2014-10-06	2014-10-06 to 2014-11-06	0.345	0.331	0.342	0.327	231	155	156
2014-08-06 to 2014-11-06	2014-11-06 to 2014-12-06	0.392	0.449	0.470	0.395	253	147	100
2014-09-06 to 2014-12-06	2014-12-06 to 2015-01-06	0.318	0.286	0.303	0.211	170	171	143

Continued on next page

Table 11 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2014-10-06 to 2015-01-06	2015-01-06 to 2015-02-06	0.376	0.316	0.320	0.306	235	205	108
2014-11-06 to 2015-02-06	2015-02-06 to 2015-03-06	0.242	0.381	0.363	0.229	231	176	69
2014-12-06 to 2015-03-06	2015-03-06 to 2015-04-06	0.374	0.370	0.364	0.358	169	133	193
2015-01-06 to 2015-04-06	2015-04-06 to 2015-05-06	0.439	0.441	0.437	0.437	205	148	166
2015-02-06 to 2015-05-06	2015-05-06 to 2015-06-06	0.425	0.408	0.406	0.405	221	145	152
2015-03-06 to 2015-06-06	2015-06-06 to 2015-07-06	0.343	0.335	0.349	0.332	250	143	100
2015-04-06 to 2015-07-06	2015-07-06 to 2015-08-06	0.344	0.325	0.321	0.322	245	151	145
2015-05-06 to 2015-08-06	2015-08-06 to 2015-09-06	0.372	0.357	0.378	0.345	220	127	147
2015-06-06 to 2015-09-06	2015-09-06 to 2015-10-06	0.298	0.270	0.281	0.274	218	131	168
2015-07-06 to 2015-10-06	2015-10-06 to 2015-11-06	0.293	0.280	0.281	0.280	221	174	148
2015-08-06 to 2015-11-06	2015-11-06 to 2015-12-06	0.248	0.257	0.258	0.221	247	140	85
2015-09-06 to 2015-12-06	2015-12-06 to 2016-01-06	0.361	0.332	0.383	0.301	227	119	161
2015-10-06 to 2016-01-06	2016-01-06 to 2016-02-06	0.442	0.400	0.392	0.391	239	84	200
2015-11-06 to 2016-02-06	2016-02-06 to 2016-03-06	0.392	0.359	0.349	0.336	221	186	68
2015-12-06 to 2016-03-06	2016-03-06 to 2016-04-06	0.352	0.380	0.368	0.296	296	131	116
2016-01-06 to 2016-04-06	2016-04-06 to 2016-05-06	0.282	0.278	0.310	0.251	202	96	219
2016-02-06 to 2016-05-06	2016-05-06 to 2016-06-06	0.295	0.267	0.256	0.258	237	138	117
2016-03-06 to 2016-06-06	2016-06-06 to 2016-07-06	0.403	0.318	0.327	0.317	249	100	169

Continued on next page

Table 11 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2016-04-06 to 2016-07-06	2016-07-06 to 2016-08-06	0.329	0.351	0.335	0.302	244	117	158
2016-05-06 to 2016-08-06	2016-08-06 to 2016-09-06	0.339	0.338	0.352	0.312	246	99	168
2016-06-06 to 2016-09-06	2016-09-06 to 2016-10-06	0.412	0.385	0.380	0.360	204	168	145
2016-07-06 to 2016-10-06	2016-10-06 to 2016-11-06	0.301	0.276	0.279	0.260	205	139	151
2016-08-06 to 2016-11-06	2016-11-06 to 2016-12-06	0.359	0.287	0.292	0.268	292	134	95
2016-09-06 to 2016-12-06	2016-12-06 to 2017-01-06	0.354	0.321	0.331	0.314	264	107	177
2016-10-06 to 2017-01-06	2017-01-06 to 2017-02-06	0.359	0.337	0.345	0.336	219	156	121
2016-11-06 to 2017-02-06	2017-02-06 to 2017-03-06	0.368	0.383	0.361	0.361	191	166	118
2016-12-06 to 2017-03-06	2017-03-06 to 2017-04-06	0.387	0.383	0.364	0.337	258	171	114
2017-01-06 to 2017-04-06	2017-04-06 to 2017-05-06	0.285	0.325	0.360	0.269	242	82	167
2017-02-06 to 2017-05-06	2017-05-06 to 2017-06-06	0.293	0.288	0.282	0.281	214	123	181
2017-03-06 to 2017-06-06	2017-06-06 to 2017-07-06	0.345	0.344	0.332	0.330	216	148	155
2017-04-06 to 2017-07-06	2017-07-06 to 2017-08-06	0.277	0.386	0.357	0.255	223	88	183
2017-05-06 to 2017-08-06	2017-08-06 to 2017-09-06	0.388	0.396	0.388	0.365	257	158	126
2017-06-06 to 2017-09-06	2017-09-06 to 2017-10-06	0.443	0.431	0.415	0.406	218	172	127
2017-07-06 to 2017-10-06	2017-10-06 to 2017-11-06	0.297	0.301	0.303	0.295	217	146	128
2017-08-06 to 2017-11-06	2017-11-06 to 2017-12-06	0.426	0.433	0.411	0.402	242	111	168
2017-09-06 to 2017-12-06	2017-12-06 to 2018-01-06	0.365	0.393	0.387	0.364	182	100	206

Continued on next page

Table 11 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2017-10-06 to 2018-01-06	2018-01-06 to 2018-02-06	0.256	0.282	0.299	0.204	203	139	178
2017-11-06 to 2018-02-06	2018-02-06 to 2018-03-06	0.335	0.321	0.325	0.319	241	117	117
2017-12-06 to 2018-03-06	2018-03-06 to 2018-04-06	0.509	0.487	0.469	0.456	290	133	117
2018-01-06 to 2018-04-06	2018-04-06 to 2018-05-06	0.335	0.313	0.303	0.303	206	180	86
2018-02-06 to 2018-05-06	2018-05-06 to 2018-06-06	0.402	0.310	0.358	0.307	211	189	140
2018-03-06 to 2018-06-06	2018-06-06 to 2018-07-06	0.316	0.325	0.336	0.316	226	148	145
2018-04-06 to 2018-07-06	2018-07-06 to 2018-08-06	0.370	0.372	0.372	0.366	202	170	122
2018-05-06 to 2018-08-06	2018-08-06 to 2018-09-06	0.310	0.308	0.295	0.298	238	160	144
2018-06-06 to 2018-09-06	2018-09-06 to 2018-10-06	0.335	0.331	0.316	0.314	197	133	163
2018-07-06 to 2018-10-06	2018-10-06 to 2018-11-06	0.277	0.291	0.300	0.236	245	135	137
2018-08-06 to 2018-11-06	2018-11-06 to 2018-12-06	0.386	0.336	0.349	0.312	209	158	154
2018-09-06 to 2018-12-06	2018-12-06 to 2019-01-06	0.387	0.384	0.384	0.371	182	107	174
2018-10-06 to 2019-01-06	2019-01-06 to 2019-02-06	0.423	0.406	0.415	0.410	265	187	94
2018-11-06 to 2019-02-06	2019-02-06 to 2019-03-06	0.284	0.305	0.320	0.270	243	156	76
2018-12-06 to 2019-03-06	2019-03-06 to 2019-04-06	0.402	0.369	0.369	0.369	258	117	143
2019-01-06 to 2019-04-06	2019-04-06 to 2019-05-06	0.357	0.325	0.322	0.323	244	129	120
2019-02-06 to 2019-05-06	2019-05-06 to 2019-06-06	0.368	0.348	0.376	0.321	254	162	127
2019-03-06 to 2019-06-06	2019-06-06 to 2019-07-06	0.307	0.273	0.270	0.270	258	123	114

Continued on next page

Table 11 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2019-04-06 to 2019-07-06	2019-07-06 to 2019-08-06	0.409	0.410	0.386	0.378	233	157	128
2019-05-06 to 2019-08-06	2019-08-06 to 2019-09-06	0.354	0.388	0.424	0.328	272	178	90
2019-06-06 to 2019-09-06	2019-09-06 to 2019-10-06	0.304	0.329	0.343	0.295	213	149	105
2019-07-06 to 2019-10-06	2019-10-06 to 2019-11-06	0.376	0.354	0.363	0.353	252	127	163
2019-08-06 to 2019-11-06	2019-11-06 to 2019-12-06	0.342	0.321	0.309	0.311	271	132	117
2019-09-06 to 2019-12-06	2019-12-06 to 2020-01-06	0.365	0.359	0.361	0.348	181	138	139
2019-10-06 to 2020-01-06	2020-01-06 to 2020-02-06	0.400	0.374	0.362	0.358	247	197	101
2019-11-06 to 2020-02-06	2020-02-06 to 2020-03-06	0.260	0.214	0.239	0.207	180	133	183
2019-12-06 to 2020-03-06	2020-03-06 to 2020-04-06	0.341	0.281	0.403	0.283	219	176	98
2020-01-06 to 2020-04-06	2020-04-06 to 2020-05-06	0.410	0.446	0.421	0.384	263	138	119
2020-02-06 to 2020-05-06	2020-05-06 to 2020-06-06	0.431	0.372	0.366	0.364	227	63	227
2020-03-06 to 2020-06-06	2020-06-06 to 2020-07-06	0.405	0.445	0.409	0.391	185	162	144
2020-04-06 to 2020-07-06	2020-07-06 to 2020-08-06	0.183	0.237	0.321	0.151	240	84	216
2020-05-06 to 2020-08-06	2020-08-06 to 2020-09-06	0.338	0.313	0.327	0.308	192	150	152
2020-06-06 to 2020-09-06	2020-09-06 to 2020-10-06	0.307	0.288	0.317	0.274	234	144	140
2020-07-06 to 2020-10-06	2020-10-06 to 2020-11-06	0.385	0.387	0.363	0.345	213	143	187
2020-08-06 to 2020-11-06	2020-11-06 to 2020-12-06	0.248	0.305	0.317	0.246	207	74	194
2020-09-06 to 2020-12-06	2020-12-06 to 2021-01-06	0.345	0.346	0.317	0.290	214	106	187

Continued on next page

Table 11 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2020-10-06 to 2021-01-06	2021-01-06 to 2021-02-06	0.300	0.368	0.340	0.293	194	212	117
2020-11-06 to 2021-02-06	2021-02-06 to 2021-03-06	0.336	0.325	0.335	0.323	165	169	142
2020-12-06 to 2021-03-06	2021-03-06 to 2021-04-06	0.338	0.399	0.368	0.329	222	155	141
2021-01-06 to 2021-04-06	2021-04-06 to 2021-05-06	0.294	0.317	0.309	0.291	232	111	174
2021-02-06 to 2021-05-06	2021-05-06 to 2021-06-06	0.359	0.386	0.372	0.361	195	133	168
2021-03-06 to 2021-06-06	2021-06-06 to 2021-07-06	0.333	0.326	0.322	0.318	245	159	110
2021-04-06 to 2021-07-06	2021-07-06 to 2021-08-06	0.424	0.368	0.357	0.345	261	170	111
2021-05-06 to 2021-08-06	2021-08-06 to 2021-09-06	0.336	0.335	0.328	0.312	267	79	145
2021-06-06 to 2021-09-06	2021-09-06 to 2021-10-06	0.323	0.302	0.296	0.287	230	200	87
2021-07-06 to 2021-10-06	2021-10-06 to 2021-11-06	0.273	0.304	0.300	0.269	278	119	120
2021-08-06 to 2021-11-06	2021-11-06 to 2021-12-06	0.352	0.233	0.267	0.232	242	161	94
2021-09-06 to 2021-12-06	2021-12-06 to 2022-01-06	0.398	0.376	0.381	0.375	215	185	143
2021-10-06 to 2022-01-06	2022-01-06 to 2022-02-06	0.376	0.381	0.374	0.372	194	165	141
2021-11-06 to 2022-02-06	2022-02-06 to 2022-03-06	0.395	0.328	0.322	0.322	217	198	58
2021-12-06 to 2022-03-06	2022-03-06 to 2022-04-06	0.252	0.311	0.258	0.243	222	152	166
2022-01-06 to 2022-04-06	2022-04-06 to 2022-05-06	0.288	0.625	0.361	0.190	222	165	131
2022-02-06 to 2022-05-06	2022-05-06 to 2022-06-06	0.341	0.336	0.336	0.321	206	139	148
2022-03-06 to 2022-06-06	2022-06-06 to 2022-07-06	0.395	0.394	0.380	0.377	207	209	101

Continued on next page

Table 11 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2022-04-06 to 2022-07-06	2022-07-06 to 2022-08-06	0.341	0.353	0.345	0.336	223	165	131
2022-05-06 to 2022-08-06	2022-08-06 to 2022-09-06	0.347	0.356	0.344	0.342	198	164	145
2022-06-06 to 2022-09-06	2022-09-06 to 2022-10-06	0.378	0.371	0.368	0.332	224	157	127
2022-07-06 to 2022-10-06	2022-10-06 to 2022-11-06	0.348	0.322	0.323	0.322	203	122	160
2022-08-06 to 2022-11-06	2022-11-06 to 2022-12-06	0.360	0.333	0.364	0.328	237	133	152
2022-09-06 to 2022-12-06	2022-12-06 to 2023-01-06	0.381	0.449	0.399	0.330	294	93	161
2022-10-06 to 2023-01-06	2023-01-06 to 2023-02-06	0.301	0.299	0.311	0.242	247	136	116
2022-11-06 to 2023-02-06	2023-02-06 to 2023-03-06	0.367	0.368	0.366	0.359	189	163	122
2022-12-06 to 2023-03-06	2023-03-06 to 2023-04-06	0.366	0.431	0.394	0.358	205	113	225
2023-01-06 to 2023-04-06	2023-04-06 to 2023-05-06	0.421	0.470	0.423	0.421	175	157	164
2023-02-06 to 2023-05-06	2023-05-06 to 2023-06-06	0.286	0.342	0.366	0.286	223	222	72
2023-03-06 to 2023-06-06	2023-06-06 to 2023-07-06	0.288	0.301	0.304	0.288	221	121	178
2023-04-06 to 2023-07-06	2023-07-06 to 2023-08-06	0.304	0.331	0.386	0.296	235	104	155
2023-05-06 to 2023-08-06	2023-08-06 to 2023-09-06	0.348	0.346	0.345	0.342	234	183	123
2023-06-06 to 2023-09-06	2023-09-06 to 2023-10-06	0.378	0.361	0.354	0.317	210	143	166
2023-07-06 to 2023-10-06	2023-10-06 to 2023-11-06	0.344	0.330	0.342	0.331	210	111	173
2023-08-06 to 2023-11-06	2023-11-06 to 2023-12-06	0.433	0.422	0.425	0.422	226	163	133
2023-09-06 to 2023-12-06	2023-12-06 to 2024-01-06	0.353	0.387	0.358	0.349	167	142	176

Continued on next page

Table 11 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2023-10-06 to 2024-01-06	2024-01-06 to 2024-02-06	0.371	0.377	0.393	0.365	241	174	105
2023-11-06 to 2024-02-06	2024-02-06 to 2024-03-06	0.437	0.385	0.384	0.379	272	72	153
2023-12-06 to 2024-03-06	2024-03-06 to 2024-04-06	0.313	0.319	0.331	0.311	191	191	139
2024-01-06 to 2024-04-06	2024-04-06 to 2024-05-06	0.462	0.407	0.359	0.315	231	126	136
2024-02-06 to 2024-05-06	2024-05-06 to 2024-06-06	0.464	0.456	0.462	0.458	247	143	153
2024-03-06 to 2024-06-06	2024-06-06 to 2024-07-06	0.336	0.336	0.336	0.336	170	165	159
2024-04-06 to 2024-07-06	2024-07-06 to 2024-08-06	0.355	0.356	0.361	0.346	233	133	152
2024-05-06 to 2024-08-06	2024-08-06 to 2024-09-06	0.336	0.364	0.317	0.321	266	116	162
2024-06-06 to 2024-09-06	2024-09-06 to 2024-10-06	0.449	0.406	0.391	0.388	220	122	130
2024-07-06 to 2024-10-06	2024-10-06 to 2024-11-06	0.363	0.348	0.331	0.325	245	167	128

Table 12: Random Forest Performance over Rolling Windows for GBP/ USD

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2008-12-26 to 2009-03-26	2009-03-26 to 2009-04-26	0.370	0.389	0.322	0.298	256	131	129
2009-01-26 to 2009-04-26	2009-04-26 to 2009-05-26	0.333	0.266	0.436	0.239	314	39	140
2009-02-26 to 2009-05-26	2009-05-26 to 2009-06-26	0.343	0.144	0.336	0.177	183	146	207
2009-03-26 to 2009-06-26	2009-06-26 to 2009-07-26	0.395	0.514	0.427	0.391	209	145	132
2009-04-26 to 2009-07-26	2009-07-26 to 2009-08-26	0.379	0.456	0.376	0.356	235	152	125
2009-05-26 to 2009-08-26	2009-08-26 to 2009-09-26	0.392	0.367	0.375	0.355	235	147	154
2009-06-26 to 2009-09-26	2009-09-26 to 2009-10-26	0.393	0.406	0.407	0.362	224	96	146
2009-07-26 to 2009-10-26	2009-10-26 to 2009-11-26	0.440	0.500	0.425	0.371	206	146	180
2009-08-26 to 2009-11-26	2009-11-26 to 2009-12-26	0.388	0.440	0.363	0.323	265	165	83
2009-09-26 to 2009-12-26	2009-12-26 to 2010-01-26	0.356	0.163	0.324	0.180	180	130	181
2009-10-26 to 2010-01-26	2010-01-26 to 2010-02-26	0.301	0.211	0.346	0.241	252	184	108
2009-11-26 to 2010-02-26	2010-02-26 to 2010-03-26	0.449	0.303	0.401	0.339	171	177	120
2009-12-26 to 2010-03-26	2010-03-26 to 2010-04-26	0.332	0.431	0.428	0.288	187	107	194
2010-01-26 to 2010-04-26	2010-04-26 to 2010-05-26	0.305	0.324	0.313	0.289	249	207	59
2010-02-26 to 2010-05-26	2010-05-26 to 2010-06-26	0.475	0.364	0.357	0.339	273	101	159
2010-03-26 to 2010-06-26	2010-06-26 to 2010-07-26	0.465	0.404	0.341	0.227	216	114	139
2010-04-26 to 2010-07-26	2010-07-26 to 2010-08-26	0.297	0.432	0.335	0.155	256	158	122

Continued on next page

Table 12 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2010-05-26 to 2010-08-26	2010-08-26 to 2010-09-26	0.412	0.137	0.333	0.194	207	149	147
2010-06-26 to 2010-09-26	2010-09-26 to 2010-10-26	0.506	0.417	0.432	0.377	238	109	145
2010-07-26 to 2010-10-26	2010-10-26 to 2010-11-26	0.337	0.253	0.322	0.265	245	161	140
2010-08-26 to 2010-11-26	2010-11-26 to 2010-12-26	0.410	0.428	0.344	0.319	213	182	102
2010-09-26 to 2010-12-26	2010-12-26 to 2011-01-26	0.441	0.299	0.358	0.277	234	100	190
2010-10-26 to 2011-01-26	2011-01-26 to 2011-02-26	0.428	0.524	0.433	0.411	208	154	183
2010-11-26 to 2011-02-26	2011-02-26 to 2011-03-26	0.385	0.396	0.367	0.324	160	173	140
2010-12-26 to 2011-03-26	2011-03-26 to 2011-04-26	0.267	0.297	0.343	0.241	259	80	152
2011-01-26 to 2011-04-26	2011-04-26 to 2011-05-26	0.399	0.206	0.305	0.238	239	139	141
2011-02-26 to 2011-05-26	2011-05-26 to 2011-06-26	0.393	0.377	0.353	0.314	202	179	136
2011-03-26 to 2011-06-26	2011-06-26 to 2011-07-26	0.518	0.315	0.398	0.349	262	116	118
2011-04-26 to 2011-07-26	2011-07-26 to 2011-08-26	0.411	0.372	0.358	0.345	227	187	128
2011-05-26 to 2011-08-26	2011-08-26 to 2011-09-26	0.294	0.434	0.321	0.276	211	187	89
2011-06-26 to 2011-09-26	2011-09-26 to 2011-10-26	0.432	0.553	0.325	0.234	236	119	161
2011-07-26 to 2011-10-26	2011-10-26 to 2011-11-26	0.452	0.151	0.333	0.208	246	176	122
2011-08-26 to 2011-11-26	2011-11-26 to 2011-12-26	0.471	0.385	0.338	0.270	232	103	141
2011-09-26 to 2011-12-26	2011-12-26 to 2012-01-26	0.437	0.289	0.381	0.327	270	118	159
2011-10-26 to 2012-01-26	2012-01-26 to 2012-02-26	0.493	0.335	0.379	0.306	247	116	158

Continued on next page

Table 12 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2011-11-26 to 2012-02-26	2012-02-26 to 2012-03-26	0.440	0.440	0.442	0.408	197	125	153
2011-12-26 to 2012-03-26	2012-03-26 to 2012-04-26	0.285	0.495	0.365	0.231	254	114	176
2012-01-26 to 2012-04-26	2012-04-26 to 2012-05-26	0.371	0.363	0.386	0.284	240	221	54
2012-02-26 to 2012-05-26	2012-05-26 to 2012-06-26	0.452	0.260	0.381	0.308	237	165	94
2012-03-26 to 2012-06-26	2012-06-26 to 2012-07-26	0.427	0.144	0.333	0.201	220	171	124
2012-04-26 to 2012-07-26	2012-07-26 to 2012-08-26	0.436	0.622	0.446	0.380	219	142	151
2012-05-26 to 2012-08-26	2012-08-26 to 2012-09-26	0.224	0.326	0.314	0.201	250	89	178
2012-06-26 to 2012-09-26	2012-09-26 to 2012-10-26	0.364	0.245	0.338	0.279	223	134	162
2012-07-26 to 2012-10-26	2012-10-26 to 2012-11-26	0.408	0.346	0.346	0.335	242	136	119
2012-08-26 to 2012-11-26	2012-11-26 to 2012-12-26	0.386	0.373	0.361	0.364	218	108	190
2012-09-26 to 2012-12-26	2012-12-26 to 2013-01-26	0.358	0.385	0.379	0.312	177	223	123
2012-10-26 to 2013-01-26	2013-01-26 to 2013-02-26	0.380	0.251	0.339	0.275	195	166	139
2012-11-26 to 2013-02-26	2013-02-26 to 2013-03-26	0.446	0.293	0.370	0.284	202	127	144
2012-12-26 to 2013-03-26	2013-03-26 to 2013-04-26	0.331	0.225	0.377	0.273	202	133	208
2013-01-26 to 2013-04-26	2013-04-26 to 2013-05-26	0.320	0.317	0.296	0.229	182	163	148
2013-02-26 to 2013-05-26	2013-05-26 to 2013-06-26	0.428	0.349	0.344	0.296	227	121	166
2013-03-26 to 2013-06-26	2013-06-26 to 2013-07-26	0.362	0.377	0.373	0.318	220	158	141
2013-04-26 to 2013-07-26	2013-07-26 to 2013-08-26	0.439	0.387	0.413	0.396	251	104	139

Continued on next page

Table 12 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2013-05-26 to 2013-08-26	2013-08-26 to 2013-09-26	0.351	0.224	0.368	0.278	206	121	211
2013-06-26 to 2013-09-26	2013-09-26 to 2013-10-26	0.349	0.325	0.360	0.311	193	153	170
2013-07-26 to 2013-10-26	2013-10-26 to 2013-11-26	0.455	0.540	0.415	0.393	201	168	130
2013-08-26 to 2013-11-26	2013-11-26 to 2013-12-26	0.236	0.140	0.334	0.139	223	118	168
2013-09-26 to 2013-12-26	2013-12-26 to 2014-01-26	0.327	0.257	0.404	0.265	220	101	174
2013-10-26 to 2014-01-26	2014-01-26 to 2014-02-26	0.356	0.324	0.326	0.286	229	130	163
2013-11-26 to 2014-02-26	2014-02-26 to 2014-03-26	0.438	0.168	0.315	0.212	220	121	132
2013-12-26 to 2014-03-26	2014-03-26 to 2014-04-26	0.406	0.382	0.428	0.329	272	125	143
2014-01-26 to 2014-04-26	2014-04-26 to 2014-05-26	0.326	0.206	0.318	0.250	185	120	167
2014-02-26 to 2014-05-26	2014-05-26 to 2014-06-26	0.434	0.425	0.399	0.373	247	122	175
2014-03-26 to 2014-06-26	2014-06-26 to 2014-07-26	0.516	0.353	0.434	0.370	294	151	72
2014-04-26 to 2014-07-26	2014-07-26 to 2014-08-26	0.347	0.531	0.417	0.304	243	171	82
2014-05-26 to 2014-08-26	2014-08-26 to 2014-09-26	0.423	0.141	0.333	0.198	230	196	118
2014-06-26 to 2014-09-26	2014-09-26 to 2014-10-26	0.347	0.300	0.305	0.295	221	136	133
2014-07-26 to 2014-10-26	2014-10-26 to 2014-11-26	0.243	0.369	0.365	0.223	264	176	83
2014-08-26 to 2014-11-26	2014-11-26 to 2014-12-26	0.540	0.365	0.374	0.326	267	147	95
2014-09-26 to 2014-12-26	2014-12-26 to 2015-01-26	0.298	0.106	0.301	0.157	157	191	128
2014-10-26 to 2015-01-26	2015-01-26 to 2015-02-26	0.224	0.283	0.299	0.211	282	120	146

Continued on next page

Table 12 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2014-11-26 to 2015-02-26	2015-02-26 to 2015-03-26	0.316	0.264	0.389	0.287	148	218	106
2014-12-26 to 2015-03-26	2015-03-26 to 2015-04-26	0.284	0.298	0.281	0.281	192	144	178
2015-01-26 to 2015-04-26	2015-04-26 to 2015-05-26	0.363	0.255	0.338	0.263	187	144	165
2015-02-26 to 2015-05-26	2015-05-26 to 2015-06-26	0.426	0.432	0.426	0.422	219	163	160
2015-03-26 to 2015-06-26	2015-06-26 to 2015-07-26	0.341	0.339	0.329	0.318	265	102	123
2015-04-26 to 2015-07-26	2015-07-26 to 2015-08-26	0.576	0.380	0.350	0.285	297	101	121
2015-05-26 to 2015-08-26	2015-08-26 to 2015-09-26	0.380	0.354	0.364	0.333	215	196	129
2015-06-26 to 2015-09-26	2015-09-26 to 2015-10-26	0.486	0.324	0.375	0.347	240	70	163
2015-07-26 to 2015-10-26	2015-10-26 to 2015-11-26	0.328	0.329	0.319	0.264	229	165	152
2015-08-26 to 2015-11-26	2015-11-26 to 2015-12-26	0.467	0.270	0.394	0.320	238	162	108
2015-09-26 to 2015-12-26	2015-12-26 to 2016-01-26	0.283	0.311	0.354	0.245	184	219	74
2015-10-26 to 2016-01-26	2016-01-26 to 2016-02-26	0.401	0.382	0.386	0.328	217	175	156
2015-11-26 to 2016-02-26	2016-02-26 to 2016-03-26	0.297	0.241	0.271	0.243	221	153	121
2015-12-26 to 2016-03-26	2016-03-26 to 2016-04-26	0.358	0.236	0.382	0.290	174	126	197
2016-01-26 to 2016-04-26	2016-04-26 to 2016-05-26	0.373	0.659	0.411	0.297	268	104	148
2016-02-26 to 2016-05-26	2016-05-26 to 2016-06-26	0.489	0.370	0.350	0.295	257	125	131
2016-03-26 to 2016-06-26	2016-06-26 to 2016-07-26	0.423	0.374	0.433	0.319	275	108	114
2016-04-26 to 2016-07-26	2016-07-26 to 2016-08-26	0.398	0.238	0.327	0.204	221	131	191

Continued on next page

Table 12 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2016-05-26 to 2016-08-26	2016-08-26 to 2016-09-26	0.507	0.473	0.467	0.465	245	131	117
2016-06-26 to 2016-09-26	2016-09-26 to 2016-10-26	0.197	0.190	0.337	0.161	258	191	70
2016-07-26 to 2016-10-26	2016-10-26 to 2016-11-26	0.256	0.267	0.271	0.243	247	109	187
2016-08-26 to 2016-11-26	2016-11-26 to 2016-12-26	0.311	0.277	0.285	0.237	177	165	134
2016-09-26 to 2016-12-26	2016-12-26 to 2017-01-26	0.378	0.332	0.332	0.304	231	151	166
2016-10-26 to 2017-01-26	2017-01-26 to 2017-02-26	0.451	0.463	0.365	0.264	216	200	105
2016-11-26 to 2017-02-26	2017-02-26 to 2017-03-26	0.278	0.345	0.242	0.220	238	115	118
2016-12-26 to 2017-03-26	2017-03-26 to 2017-04-26	0.323	0.204	0.300	0.235	247	88	185
2017-01-26 to 2017-04-26	2017-04-26 to 2017-05-26	0.451	0.151	0.333	0.207	234	123	162
2017-02-26 to 2017-05-26	2017-05-26 to 2017-06-26	0.552	0.362	0.375	0.320	257	82	152
2017-03-26 to 2017-06-26	2017-06-26 to 2017-07-26	0.417	0.266	0.327	0.287	233	96	189
2017-04-26 to 2017-07-26	2017-07-26 to 2017-08-26	0.484	0.281	0.369	0.315	277	140	124
2017-05-26 to 2017-08-26	2017-08-26 to 2017-09-26	0.467	0.330	0.366	0.294	220	112	165
2017-06-26 to 2017-09-26	2017-09-26 to 2017-10-26	0.504	0.169	0.333	0.225	262	171	87
2017-07-26 to 2017-10-26	2017-10-26 to 2017-11-26	0.446	0.423	0.388	0.382	238	103	179
2017-08-26 to 2017-11-26	2017-11-26 to 2017-12-26	0.554	0.343	0.391	0.345	264	114	109
2017-09-26 to 2017-12-26	2017-12-26 to 2018-01-26	0.289	0.222	0.410	0.246	181	104	238
2017-10-26 to 2018-01-26	2018-01-26 to 2018-02-26	0.220	0.184	0.329	0.123	190	198	111

Continued on next page

Table 12 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2017-11-26 to 2018-02-26	2018-02-26 to 2018-03-26	0.371	0.322	0.321	0.322	210	89	175
2017-12-26 to 2018-03-26	2018-03-26 to 2018-04-26	0.333	0.370	0.319	0.306	244	160	139
2018-01-26 to 2018-04-26	2018-04-26 to 2018-05-26	0.342	0.217	0.349	0.252	243	208	66
2018-02-26 to 2018-05-26	2018-05-26 to 2018-06-26	0.416	0.306	0.402	0.318	247	128	122
2018-03-26 to 2018-06-26	2018-06-26 to 2018-07-26	0.381	0.407	0.375	0.285	177	172	171
2018-04-26 to 2018-07-26	2018-07-26 to 2018-08-26	0.277	0.321	0.363	0.219	260	153	104
2018-05-26 to 2018-08-26	2018-08-26 to 2018-09-26	0.374	0.232	0.373	0.285	206	113	202
2018-06-26 to 2018-09-26	2018-09-26 to 2018-10-26	0.308	0.287	0.301	0.291	188	189	143
2018-07-26 to 2018-10-26	2018-10-26 to 2018-11-26	0.423	0.239	0.340	0.265	219	145	130
2018-08-26 to 2018-11-26	2018-11-26 to 2018-12-26	0.371	0.258	0.353	0.224	208	125	176
2018-09-26 to 2018-12-26	2018-12-26 to 2019-01-26	0.414	0.390	0.373	0.358	224	121	177
2018-10-26 to 2019-01-26	2019-01-26 to 2019-02-26	0.353	0.284	0.276	0.247	247	146	106
2018-11-26 to 2019-02-26	2019-02-26 to 2019-03-26	0.496	0.165	0.333	0.221	235	148	91
2018-12-26 to 2019-03-26	2019-03-26 to 2019-04-26	0.472	0.285	0.337	0.288	278	140	126
2019-01-26 to 2019-04-26	2019-04-26 to 2019-05-26	0.402	0.507	0.404	0.318	194	210	89
2019-02-26 to 2019-05-26	2019-05-26 to 2019-06-26	0.354	0.344	0.345	0.339	222	159	139
2019-03-26 to 2019-06-26	2019-06-26 to 2019-07-26	0.361	0.411	0.397	0.318	174	196	148
2019-04-26 to 2019-07-26	2019-07-26 to 2019-08-26	0.249	0.083	0.333	0.133	239	132	123

Continued on next page

Table 12 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2019-05-26 to 2019-08-26	2019-08-26 to 2019-09-26	0.381	0.349	0.344	0.336	238	164	141
2019-06-26 to 2019-09-26	2019-09-26 to 2019-10-26	0.381	0.268	0.287	0.240	244	115	158
2019-07-26 to 2019-10-26	2019-10-26 to 2019-11-26	0.455	0.225	0.331	0.223	231	125	143
2019-08-26 to 2019-11-26	2019-11-26 to 2019-12-26	0.436	0.303	0.403	0.330	247	129	131
2019-09-26 to 2019-12-26	2019-12-26 to 2020-01-26	0.476	0.461	0.394	0.319	215	131	152
2019-10-26 to 2020-01-26	2020-01-26 to 2020-02-26	0.416	0.693	0.339	0.209	215	179	128
2019-11-26 to 2020-02-26	2020-02-26 to 2020-03-26	0.345	0.140	0.300	0.184	192	174	132
2019-12-26 to 2020-03-26	2020-03-26 to 2020-04-26	0.403	0.334	0.375	0.336	253	113	150
2020-01-26 to 2020-04-26	2020-04-26 to 2020-05-26	0.387	0.241	0.313	0.258	224	148	124
2020-02-26 to 2020-05-26	2020-05-26 to 2020-06-26	0.346	0.262	0.306	0.239	215	157	172
2020-03-26 to 2020-06-26	2020-06-26 to 2020-07-26	0.323	0.278	0.334	0.266	219	108	166
2020-04-26 to 2020-07-26	2020-07-26 to 2020-08-26	0.190	0.064	0.333	0.108	231	99	191
2020-05-26 to 2020-08-26	2020-08-26 to 2020-09-26	0.312	0.324	0.334	0.312	243	171	127
2020-06-26 to 2020-09-26	2020-09-26 to 2020-10-26	0.478	0.364	0.329	0.245	235	124	114
2020-07-26 to 2020-10-26	2020-10-26 to 2020-11-26	0.455	0.183	0.304	0.218	275	92	180
2020-08-26 to 2020-11-26	2020-11-26 to 2020-12-26	0.359	0.243	0.373	0.271	181	157	169
2020-09-26 to 2020-12-26	2020-12-26 to 2021-01-26	0.247	0.248	0.338	0.140	202	115	160
2020-10-26 to 2021-01-26	2021-01-26 to 2021-02-26	0.283	0.221	0.327	0.220	246	131	170

Continued on next page

Table 12 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2020-11-26 to 2021-02-26	2021-02-26 to 2021-03-26	0.434	0.237	0.318	0.271	244	116	115
2020-12-26 to 2021-03-26	2021-03-26 to 2021-04-26	0.336	0.225	0.337	0.264	219	151	124
2021-01-26 to 2021-04-26	2021-04-26 to 2021-05-26	0.490	0.394	0.385	0.323	277	77	166
2021-02-26 to 2021-05-26	2021-05-26 to 2021-06-26	0.379	0.270	0.328	0.220	208	209	124
2021-03-26 to 2021-06-26	2021-06-26 to 2021-07-26	0.374	0.309	0.317	0.223	189	154	130
2021-04-26 to 2021-07-26	2021-07-26 to 2021-08-26	0.434	0.391	0.383	0.380	269	150	125
2021-05-26 to 2021-08-26	2021-08-26 to 2021-09-26	0.385	0.393	0.349	0.302	202	163	152
2021-06-26 to 2021-09-26	2021-09-26 to 2021-10-26	0.360	0.238	0.323	0.257	200	117	180
2021-07-26 to 2021-10-26	2021-10-26 to 2021-11-26	0.231	0.433	0.329	0.212	271	190	85
2021-08-26 to 2021-11-26	2021-11-26 to 2021-12-26	0.458	0.306	0.395	0.333	264	124	110
2021-09-26 to 2021-12-26	2021-12-26 to 2022-01-26	0.398	0.200	0.328	0.207	214	155	156
2021-10-26 to 2022-01-26	2022-01-26 to 2022-02-26	0.396	0.407	0.381	0.361	250	154	142
2021-11-26 to 2022-02-26	2022-02-26 to 2022-03-26	0.571	0.190	0.333	0.242	270	130	73
2021-12-26 to 2022-03-26	2022-03-26 to 2022-04-26	0.461	0.259	0.364	0.301	275	139	83
2022-01-26 to 2022-04-26	2022-04-26 to 2022-05-26	0.484	0.490	0.366	0.272	236	144	136
2022-02-26 to 2022-05-26	2022-05-26 to 2022-06-26	0.395	0.344	0.339	0.305	196	229	92
2022-03-26 to 2022-06-26	2022-06-26 to 2022-07-26	0.328	0.231	0.295	0.210	187	199	108
2022-04-26 to 2022-07-26	2022-07-26 to 2022-08-26	0.371	0.341	0.338	0.338	252	187	105

Continued on next page

Table 12 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2022-05-26 to 2022-08-26	2022-08-26 to 2022-09-26	0.350	0.180	0.319	0.228	194	199	101
2022-06-26 to 2022-09-26	2022-09-26 to 2022-10-26	0.194	0.175	0.326	0.118	214	100	202
2022-07-26 to 2022-10-26	2022-10-26 to 2022-11-26	0.384	0.209	0.308	0.244	251	119	174
2022-08-26 to 2022-11-26	2022-11-26 to 2022-12-26	0.370	0.268	0.351	0.279	255	114	107
2022-09-26 to 2022-12-26	2022-12-26 to 2023-01-26	0.552	0.280	0.362	0.311	326	73	148
2022-10-26 to 2023-01-26	2023-01-26 to 2023-02-26	0.489	0.448	0.389	0.350	257	179	85
2022-11-26 to 2023-02-26	2023-02-26 to 2023-03-26	0.401	0.287	0.332	0.209	191	107	176
2022-12-26 to 2023-03-26	2023-03-26 to 2023-04-26	0.417	0.558	0.383	0.327	220	124	176
2023-01-26 to 2023-04-26	2023-04-26 to 2023-05-26	0.408	0.412	0.388	0.359	206	171	143
2023-02-26 to 2023-05-26	2023-05-26 to 2023-06-26	0.372	0.375	0.369	0.349	218	102	174
2023-03-26 to 2023-06-26	2023-06-26 to 2023-07-26	0.409	0.345	0.334	0.273	223	125	170
2023-04-26 to 2023-07-26	2023-07-26 to 2023-08-26	0.425	0.276	0.343	0.253	232	164	143
2023-05-26 to 2023-08-26	2023-08-26 to 2023-09-26	0.257	0.181	0.354	0.217	205	222	68
2023-06-26 to 2023-09-26	2023-09-26 to 2023-10-26	0.315	0.293	0.346	0.197	235	152	133
2023-07-26 to 2023-10-26	2023-10-26 to 2023-11-26	0.369	0.393	0.416	0.351	197	111	210
2023-08-26 to 2023-11-26	2023-11-26 to 2023-12-26	0.415	0.379	0.364	0.317	215	141	128
2023-09-26 to 2023-12-26	2023-12-26 to 2024-01-26	0.403	0.345	0.386	0.318	263	145	116
2023-10-26 to 2024-01-26	2024-01-26 to 2024-02-26	0.465	0.420	0.376	0.332	232	131	134

Continued on next page

Table 12 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2023-11-26 to 2024-02-26	2024-02-26 to 2024-03-26	0.415	0.422	0.400	0.391	204	156	136
2023-12-26 to 2024-03-26	2024-03-26 to 2024-04-26	0.382	0.350	0.339	0.311	251	144	149
2024-01-26 to 2024-04-26	2024-04-26 to 2024-05-26	0.359	0.232	0.275	0.219	224	121	148
2024-02-26 to 2024-05-26	2024-05-26 to 2024-06-26	0.413	0.371	0.366	0.353	230	167	123
2024-03-26 to 2024-06-26	2024-06-26 to 2024-07-26	0.387	0.427	0.340	0.289	210	114	195
2024-04-26 to 2024-07-26	2024-07-26 to 2024-08-26	0.395	0.339	0.317	0.263	219	96	176
2024-05-26 to 2024-08-26	2024-08-26 to 2024-09-26	0.355	0.417	0.355	0.334	261	103	180
2024-06-26 to 2024-09-26	2024-09-26 to 2024-10-26	0.472	0.329	0.327	0.318	301	142	74
2024-07-26 to 2024-10-26	2024-10-26 to 2024-11-26	0.282	0.284	0.304	0.173	154	210	136
2024-08-26 to 2024-11-26	2024-11-26 to 2024-12-26	0.261	0.443	0.370	0.227	207	102	200

Table 13: MLP Performance over Rolling Windows for GBP/ USD

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2008-12-26 to 2009-03-26	2009-03-26 to 2009-04-26	0.355	0.348	0.349	0.344	256	131	129
2009-01-26 to 2009-04-26	2009-04-26 to 2009-05-26	0.357	0.611	0.477	0.271	314	39	140
2009-02-26 to 2009-05-26	2009-05-26 to 2009-06-26	0.317	0.298	0.307	0.265	183	146	207
2009-03-26 to 2009-06-26	2009-06-26 to 2009-07-26	0.438	0.460	0.454	0.440	209	145	132
2009-04-26 to 2009-07-26	2009-07-26 to 2009-08-26	0.400	0.415	0.399	0.388	235	152	125
2009-05-26 to 2009-08-26	2009-08-26 to 2009-09-26	0.265	0.265	0.265	0.261	235	147	154
2009-06-26 to 2009-09-26	2009-09-26 to 2009-10-26	0.361	0.389	0.357	0.308	224	96	146
2009-07-26 to 2009-10-26	2009-10-26 to 2009-11-26	0.389	0.397	0.393	0.386	206	146	180
2009-08-26 to 2009-11-26	2009-11-26 to 2009-12-26	0.331	0.301	0.334	0.289	265	165	83
2009-09-26 to 2009-12-26	2009-12-26 to 2010-01-26	0.330	0.341	0.338	0.329	180	130	181
2009-10-26 to 2010-01-26	2010-01-26 to 2010-02-26	0.301	0.420	0.362	0.274	252	184	108
2009-11-26 to 2010-02-26	2010-02-26 to 2010-03-26	0.397	0.268	0.357	0.306	171	177	120
2009-12-26 to 2010-03-26	2010-03-26 to 2010-04-26	0.307	0.326	0.350	0.304	187	107	194
2010-01-26 to 2010-04-26	2010-04-26 to 2010-05-26	0.322	0.338	0.340	0.296	249	207	59
2010-02-26 to 2010-05-26	2010-05-26 to 2010-06-26	0.366	0.383	0.383	0.358	273	101	159
2010-03-26 to 2010-06-26	2010-06-26 to 2010-07-26	0.409	0.382	0.381	0.381	216	114	139
2010-04-26 to 2010-07-26	2010-07-26 to 2010-08-26	0.326	0.467	0.351	0.211	256	158	122

Continued on next page

Table 13 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2010-05-26 to 2010-08-26	2010-08-26 to 2010-09-26	0.324	0.310	0.310	0.310	207	149	147
2010-06-26 to 2010-09-26	2010-09-26 to 2010-10-26	0.433	0.436	0.432	0.414	238	109	145
2010-07-26 to 2010-10-26	2010-10-26 to 2010-11-26	0.416	0.373	0.397	0.361	245	161	140
2010-08-26 to 2010-11-26	2010-11-26 to 2010-12-26	0.382	0.404	0.378	0.374	213	182	102
2010-09-26 to 2010-12-26	2010-12-26 to 2011-01-26	0.378	0.345	0.344	0.332	234	100	190
2010-10-26 to 2011-01-26	2011-01-26 to 2011-02-26	0.323	0.329	0.332	0.308	208	154	183
2010-11-26 to 2011-02-26	2011-02-26 to 2011-03-26	0.372	0.370	0.363	0.360	160	173	140
2010-12-26 to 2011-03-26	2011-03-26 to 2011-04-26	0.273	0.296	0.276	0.258	259	80	152
2011-01-26 to 2011-04-26	2011-04-26 to 2011-05-26	0.366	0.422	0.376	0.355	239	139	141
2011-02-26 to 2011-05-26	2011-05-26 to 2011-06-26	0.342	0.338	0.337	0.337	202	179	136
2011-03-26 to 2011-06-26	2011-06-26 to 2011-07-26	0.488	0.411	0.414	0.402	262	116	118
2011-04-26 to 2011-07-26	2011-07-26 to 2011-08-26	0.297	0.257	0.265	0.259	227	187	128
2011-05-26 to 2011-08-26	2011-08-26 to 2011-09-26	0.283	0.346	0.366	0.278	211	187	89
2011-06-26 to 2011-09-26	2011-09-26 to 2011-10-26	0.378	0.360	0.341	0.338	236	119	161
2011-07-26 to 2011-10-26	2011-10-26 to 2011-11-26	0.360	0.307	0.313	0.306	246	176	122
2011-08-26 to 2011-11-26	2011-11-26 to 2011-12-26	0.374	0.382	0.400	0.365	232	103	141
2011-09-26 to 2011-12-26	2011-12-26 to 2012-01-26	0.388	0.363	0.367	0.363	270	118	159
2011-10-26 to 2012-01-26	2012-01-26 to 2012-02-26	0.355	0.340	0.355	0.331	247	116	158

Continued on next page

Table 13 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2011-11-26 to 2012-02-26	2012-02-26 to 2012-03-26	0.383	0.384	0.381	0.380	197	125	153
2011-12-26 to 2012-03-26	2012-03-26 to 2012-04-26	0.305	0.389	0.366	0.289	254	114	176
2012-01-26 to 2012-04-26	2012-04-26 to 2012-05-26	0.311	0.303	0.285	0.269	240	221	54
2012-02-26 to 2012-05-26	2012-05-26 to 2012-06-26	0.333	0.341	0.359	0.291	237	165	94
2012-03-26 to 2012-06-26	2012-06-26 to 2012-07-26	0.396	0.389	0.385	0.387	220	171	124
2012-04-26 to 2012-07-26	2012-07-26 to 2012-08-26	0.354	0.312	0.340	0.299	219	142	151
2012-05-26 to 2012-08-26	2012-08-26 to 2012-09-26	0.213	0.284	0.303	0.194	250	89	178
2012-06-26 to 2012-09-26	2012-09-26 to 2012-10-26	0.360	0.448	0.365	0.326	223	134	162
2012-07-26 to 2012-10-26	2012-10-26 to 2012-11-26	0.396	0.384	0.380	0.375	242	136	119
2012-08-26 to 2012-11-26	2012-11-26 to 2012-12-26	0.349	0.357	0.364	0.346	218	108	190
2012-09-26 to 2012-12-26	2012-12-26 to 2013-01-26	0.455	0.491	0.464	0.431	177	223	123
2012-10-26 to 2013-01-26	2013-01-26 to 2013-02-26	0.292	0.290	0.287	0.283	195	166	139
2012-11-26 to 2013-02-26	2013-02-26 to 2013-03-26	0.334	0.417	0.370	0.255	202	127	144
2012-12-26 to 2013-03-26	2013-03-26 to 2013-04-26	0.293	0.310	0.336	0.276	202	133	208
2013-01-26 to 2013-04-26	2013-04-26 to 2013-05-26	0.302	0.302	0.304	0.303	182	163	148
2013-02-26 to 2013-05-26	2013-05-26 to 2013-06-26	0.313	0.360	0.311	0.292	227	121	166
2013-03-26 to 2013-06-26	2013-06-26 to 2013-07-26	0.337	0.363	0.366	0.331	220	158	141
2013-04-26 to 2013-07-26	2013-07-26 to 2013-08-26	0.342	0.305	0.299	0.300	251	104	139

Continued on next page

Table 13 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2013-05-26 to 2013-08-26	2013-08-26 to 2013-09-26	0.359	0.364	0.352	0.340	206	121	211
2013-06-26 to 2013-09-26	2013-09-26 to 2013-10-26	0.349	0.365	0.358	0.339	193	153	170
2013-07-26 to 2013-10-26	2013-10-26 to 2013-11-26	0.407	0.422	0.426	0.406	201	168	130
2013-08-26 to 2013-11-26	2013-11-26 to 2013-12-26	0.312	0.510	0.364	0.262	223	118	168
2013-09-26 to 2013-12-26	2013-12-26 to 2014-01-26	0.303	0.408	0.382	0.270	220	101	174
2013-10-26 to 2014-01-26	2014-01-26 to 2014-02-26	0.287	0.301	0.308	0.289	229	130	163
2013-11-26 to 2014-02-26	2014-02-26 to 2014-03-26	0.359	0.304	0.335	0.305	220	121	132
2013-12-26 to 2014-03-26	2014-03-26 to 2014-04-26	0.344	0.360	0.342	0.334	272	125	143
2014-01-26 to 2014-04-26	2014-04-26 to 2014-05-26	0.328	0.469	0.334	0.279	185	120	167
2014-02-26 to 2014-05-26	2014-05-26 to 2014-06-26	0.415	0.433	0.400	0.399	247	122	175
2014-03-26 to 2014-06-26	2014-06-26 to 2014-07-26	0.391	0.357	0.364	0.342	294	151	72
2014-04-26 to 2014-07-26	2014-07-26 to 2014-08-26	0.393	0.366	0.396	0.354	243	171	82
2014-05-26 to 2014-08-26	2014-08-26 to 2014-09-26	0.426	0.572	0.358	0.270	230	196	118
2014-06-26 to 2014-09-26	2014-09-26 to 2014-10-26	0.402	0.375	0.381	0.378	221	136	133
2014-07-26 to 2014-10-26	2014-10-26 to 2014-11-26	0.289	0.340	0.308	0.282	264	176	83
2014-08-26 to 2014-11-26	2014-11-26 to 2014-12-26	0.430	0.322	0.330	0.321	267	147	95
2014-09-26 to 2014-12-26	2014-12-26 to 2015-01-26	0.311	0.196	0.330	0.245	157	191	128
2014-10-26 to 2015-01-26	2015-01-26 to 2015-02-26	0.292	0.330	0.333	0.281	282	120	146

Continued on next page

Table 13 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2014-11-26 to 2015-02-26	2015-02-26 to 2015-03-26	0.246	0.300	0.272	0.246	148	218	106
2014-12-26 to 2015-03-26	2015-03-26 to 2015-04-26	0.311	0.297	0.317	0.299	192	144	178
2015-01-26 to 2015-04-26	2015-04-26 to 2015-05-26	0.377	0.373	0.366	0.361	187	144	165
2015-02-26 to 2015-05-26	2015-05-26 to 2015-06-26	0.325	0.312	0.310	0.309	219	163	160
2015-03-26 to 2015-06-26	2015-06-26 to 2015-07-26	0.288	0.282	0.245	0.258	265	102	123
2015-04-26 to 2015-07-26	2015-07-26 to 2015-08-26	0.534	0.532	0.562	0.531	297	101	121
2015-05-26 to 2015-08-26	2015-08-26 to 2015-09-26	0.361	0.380	0.352	0.350	215	196	129
2015-06-26 to 2015-09-26	2015-09-26 to 2015-10-26	0.410	0.394	0.402	0.383	240	70	163
2015-07-26 to 2015-10-26	2015-10-26 to 2015-11-26	0.299	0.335	0.318	0.267	229	165	152
2015-08-26 to 2015-11-26	2015-11-26 to 2015-12-26	0.437	0.511	0.391	0.332	238	162	108
2015-09-26 to 2015-12-26	2015-12-26 to 2016-01-26	0.350	0.337	0.301	0.234	184	219	74
2015-10-26 to 2016-01-26	2016-01-26 to 2016-02-26	0.376	0.336	0.345	0.323	217	175	156
2015-11-26 to 2016-02-26	2016-02-26 to 2016-03-26	0.370	0.366	0.372	0.367	221	153	121
2015-12-26 to 2016-03-26	2016-03-26 to 2016-04-26	0.272	0.289	0.294	0.257	174	126	197
2016-01-26 to 2016-04-26	2016-04-26 to 2016-05-26	0.383	0.412	0.406	0.353	268	104	148
2016-02-26 to 2016-05-26	2016-05-26 to 2016-06-26	0.421	0.320	0.331	0.317	257	125	131
2016-03-26 to 2016-06-26	2016-06-26 to 2016-07-26	0.310	0.329	0.354	0.229	275	108	114
2016-04-26 to 2016-07-26	2016-07-26 to 2016-08-26	0.330	0.291	0.323	0.292	221	131	191

Continued on next page

Table 13 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2016-05-26 to 2016-08-26	2016-08-26 to 2016-09-26	0.345	0.319	0.334	0.309	245	131	117
2016-06-26 to 2016-09-26	2016-09-26 to 2016-10-26	0.233	0.288	0.362	0.212	258	191	70
2016-07-26 to 2016-10-26	2016-10-26 to 2016-11-26	0.241	0.235	0.257	0.209	247	109	187
2016-08-26 to 2016-11-26	2016-11-26 to 2016-12-26	0.282	0.271	0.273	0.271	177	165	134
2016-09-26 to 2016-12-26	2016-12-26 to 2017-01-26	0.343	0.330	0.329	0.329	231	151	166
2016-10-26 to 2017-01-26	2017-01-26 to 2017-02-26	0.351	0.343	0.333	0.332	216	200	105
2016-11-26 to 2017-02-26	2017-02-26 to 2017-03-26	0.295	0.299	0.294	0.287	238	115	118
2016-12-26 to 2017-03-26	2017-03-26 to 2017-04-26	0.233	0.273	0.283	0.218	247	88	185
2017-01-26 to 2017-04-26	2017-04-26 to 2017-05-26	0.432	0.233	0.330	0.252	234	123	162
2017-02-26 to 2017-05-26	2017-05-26 to 2017-06-26	0.358	0.293	0.280	0.286	257	82	152
2017-03-26 to 2017-06-26	2017-06-26 to 2017-07-26	0.359	0.241	0.313	0.245	233	96	189
2017-04-26 to 2017-07-26	2017-07-26 to 2017-08-26	0.473	0.429	0.409	0.371	277	140	124
2017-05-26 to 2017-08-26	2017-08-26 to 2017-09-26	0.469	0.402	0.400	0.382	220	112	165
2017-06-26 to 2017-09-26	2017-09-26 to 2017-10-26	0.408	0.358	0.355	0.355	262	171	87
2017-07-26 to 2017-10-26	2017-10-26 to 2017-11-26	0.433	0.421	0.425	0.418	238	103	179
2017-08-26 to 2017-11-26	2017-11-26 to 2017-12-26	0.458	0.394	0.421	0.387	264	114	109
2017-09-26 to 2017-12-26	2017-12-26 to 2018-01-26	0.293	0.464	0.417	0.259	181	104	238
2017-10-26 to 2018-01-26	2018-01-26 to 2018-02-26	0.255	0.441	0.342	0.207	190	198	111

Continued on next page

Table 13 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2017-11-26 to 2018-02-26	2018-02-26 to 2018-03-26	0.297	0.333	0.317	0.296	210	89	175
2017-12-26 to 2018-03-26	2018-03-26 to 2018-04-26	0.363	0.383	0.408	0.346	244	160	139
2018-01-26 to 2018-04-26	2018-04-26 to 2018-05-26	0.335	0.414	0.455	0.284	243	208	66
2018-02-26 to 2018-05-26	2018-05-26 to 2018-06-26	0.443	0.425	0.441	0.403	247	128	122
2018-03-26 to 2018-06-26	2018-06-26 to 2018-07-26	0.398	0.416	0.394	0.355	177	172	171
2018-04-26 to 2018-07-26	2018-07-26 to 2018-08-26	0.327	0.334	0.333	0.274	260	153	104
2018-05-26 to 2018-08-26	2018-08-26 to 2018-09-26	0.340	0.350	0.343	0.318	206	113	202
2018-06-26 to 2018-09-26	2018-09-26 to 2018-10-26	0.319	0.319	0.323	0.318	188	189	143
2018-07-26 to 2018-10-26	2018-10-26 to 2018-11-26	0.360	0.338	0.347	0.335	219	145	130
2018-08-26 to 2018-11-26	2018-11-26 to 2018-12-26	0.495	0.588	0.465	0.439	208	125	176
2018-09-26 to 2018-12-26	2018-12-26 to 2019-01-26	0.420	0.422	0.424	0.414	224	121	177
2018-10-26 to 2019-01-26	2019-01-26 to 2019-02-26	0.331	0.327	0.351	0.332	247	146	106
2018-11-26 to 2019-02-26	2019-02-26 to 2019-03-26	0.487	0.397	0.365	0.336	235	148	91
2018-12-26 to 2019-03-26	2019-03-26 to 2019-04-26	0.311	0.311	0.297	0.286	278	140	126
2019-01-26 to 2019-04-26	2019-04-26 to 2019-05-26	0.310	0.320	0.335	0.283	194	210	89
2019-02-26 to 2019-05-26	2019-05-26 to 2019-06-26	0.404	0.434	0.430	0.389	222	159	139
2019-03-26 to 2019-06-26	2019-06-26 to 2019-07-26	0.342	0.348	0.361	0.322	174	196	148
2019-04-26 to 2019-07-26	2019-07-26 to 2019-08-26	0.462	0.645	0.438	0.380	239	132	123

Continued on next page

Table 13 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2019-05-26 to 2019-08-26	2019-08-26 to 2019-09-26	0.394	0.390	0.387	0.383	238	164	141
2019-06-26 to 2019-09-26	2019-09-26 to 2019-10-26	0.348	0.322	0.330	0.312	244	115	158
2019-07-26 to 2019-10-26	2019-10-26 to 2019-11-26	0.321	0.228	0.318	0.249	231	125	143
2019-08-26 to 2019-11-26	2019-11-26 to 2019-12-26	0.410	0.401	0.412	0.357	247	129	131
2019-09-26 to 2019-12-26	2019-12-26 to 2020-01-26	0.331	0.330	0.343	0.318	215	131	152
2019-10-26 to 2020-01-26	2020-01-26 to 2020-02-26	0.412	0.402	0.401	0.400	215	179	128
2019-11-26 to 2020-02-26	2020-02-26 to 2020-03-26	0.239	0.216	0.217	0.207	192	174	132
2019-12-26 to 2020-03-26	2020-03-26 to 2020-04-26	0.351	0.319	0.344	0.290	253	113	150
2020-01-26 to 2020-04-26	2020-04-26 to 2020-05-26	0.399	0.469	0.440	0.391	224	148	124
2020-02-26 to 2020-05-26	2020-05-26 to 2020-06-26	0.351	0.346	0.348	0.345	215	157	172
2020-03-26 to 2020-06-26	2020-06-26 to 2020-07-26	0.312	0.307	0.316	0.304	219	108	166
2020-04-26 to 2020-07-26	2020-07-26 to 2020-08-26	0.263	0.310	0.337	0.250	231	99	191
2020-05-26 to 2020-08-26	2020-08-26 to 2020-09-26	0.375	0.386	0.383	0.371	243	171	127
2020-06-26 to 2020-09-26	2020-09-26 to 2020-10-26	0.376	0.359	0.364	0.356	235	124	114
2020-07-26 to 2020-10-26	2020-10-26 to 2020-11-26	0.448	0.413	0.412	0.406	275	92	180
2020-08-26 to 2020-11-26	2020-11-26 to 2020-12-26	0.363	0.383	0.366	0.325	181	157	169
2020-09-26 to 2020-12-26	2020-12-26 to 2021-01-26	0.314	0.495	0.389	0.250	202	115	160
2020-10-26 to 2021-01-26	2021-01-26 to 2021-02-26	0.291	0.362	0.325	0.272	246	131	170

Continued on next page

Table 13 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2020-11-26 to 2021-02-26	2021-02-26 to 2021-03-26	0.383	0.328	0.332	0.328	244	116	115
2020-12-26 to 2021-03-26	2021-03-26 to 2021-04-26	0.356	0.349	0.351	0.349	219	151	124
2021-01-26 to 2021-04-26	2021-04-26 to 2021-05-26	0.363	0.336	0.343	0.325	277	77	166
2021-02-26 to 2021-05-26	2021-05-26 to 2021-06-26	0.373	0.386	0.373	0.362	208	209	124
2021-03-26 to 2021-06-26	2021-06-26 to 2021-07-26	0.317	0.316	0.309	0.299	189	154	130
2021-04-26 to 2021-07-26	2021-07-26 to 2021-08-26	0.340	0.358	0.377	0.334	269	150	125
2021-05-26 to 2021-08-26	2021-08-26 to 2021-09-26	0.261	0.262	0.266	0.262	202	163	152
2021-06-26 to 2021-09-26	2021-09-26 to 2021-10-26	0.356	0.306	0.327	0.298	200	117	180
2021-07-26 to 2021-10-26	2021-10-26 to 2021-11-26	0.299	0.430	0.376	0.294	271	190	85
2021-08-26 to 2021-11-26	2021-11-26 to 2021-12-26	0.478	0.256	0.313	0.252	264	124	110
2021-09-26 to 2021-12-26	2021-12-26 to 2022-01-26	0.394	0.370	0.370	0.369	214	155	156
2021-10-26 to 2022-01-26	2022-01-26 to 2022-02-26	0.374	0.360	0.355	0.353	250	154	142
2021-11-26 to 2022-02-26	2022-02-26 to 2022-03-26	0.414	0.370	0.351	0.311	270	130	73
2021-12-26 to 2022-03-26	2022-03-26 to 2022-04-26	0.410	0.413	0.406	0.323	275	139	83
2022-01-26 to 2022-04-26	2022-04-26 to 2022-05-26	0.322	0.346	0.368	0.245	236	144	136
2022-02-26 to 2022-05-26	2022-05-26 to 2022-06-26	0.406	0.402	0.428	0.402	196	229	92
2022-03-26 to 2022-06-26	2022-06-26 to 2022-07-26	0.379	0.409	0.380	0.375	187	199	108
2022-04-26 to 2022-07-26	2022-07-26 to 2022-08-26	0.327	0.288	0.293	0.282	252	187	105

Continued on next page

Table 13 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2022-05-26 to 2022-08-26	2022-08-26 to 2022-09-26	0.366	0.364	0.369	0.285	194	199	101
2022-06-26 to 2022-09-26	2022-09-26 to 2022-10-26	0.242	0.445	0.356	0.206	214	100	202
2022-07-26 to 2022-10-26	2022-10-26 to 2022-11-26	0.298	0.291	0.298	0.286	251	119	174
2022-08-26 to 2022-11-26	2022-11-26 to 2022-12-26	0.504	0.282	0.361	0.313	255	114	107
2022-09-26 to 2022-12-26	2022-12-26 to 2023-01-26	0.331	0.313	0.328	0.303	326	73	148
2022-10-26 to 2023-01-26	2023-01-26 to 2023-02-26	0.443	0.427	0.432	0.403	257	179	85
2022-11-26 to 2023-02-26	2023-02-26 to 2023-03-26	0.445	0.412	0.408	0.404	191	107	176
2022-12-26 to 2023-03-26	2023-03-26 to 2023-04-26	0.340	0.432	0.352	0.316	220	124	176
2023-01-26 to 2023-04-26	2023-04-26 to 2023-05-26	0.406	0.431	0.392	0.398	206	171	143
2023-02-26 to 2023-05-26	2023-05-26 to 2023-06-26	0.306	0.311	0.335	0.306	218	102	174
2023-03-26 to 2023-06-26	2023-06-26 to 2023-07-26	0.392	0.349	0.339	0.318	223	125	170
2023-04-26 to 2023-07-26	2023-07-26 to 2023-08-26	0.367	0.354	0.346	0.336	232	164	143
2023-05-26 to 2023-08-26	2023-08-26 to 2023-09-26	0.303	0.403	0.385	0.284	205	222	68
2023-06-26 to 2023-09-26	2023-09-26 to 2023-10-26	0.294	0.308	0.322	0.193	235	152	133
2023-07-26 to 2023-10-26	2023-10-26 to 2023-11-26	0.309	0.343	0.357	0.306	197	111	210
2023-08-26 to 2023-11-26	2023-11-26 to 2023-12-26	0.366	0.339	0.337	0.332	215	141	128
2023-09-26 to 2023-12-26	2023-12-26 to 2024-01-26	0.399	0.406	0.418	0.395	263	145	116
2023-10-26 to 2024-01-26	2024-01-26 to 2024-02-26	0.445	0.429	0.440	0.398	232	131	134

Continued on next page

Table 13 – continued from previous page

Train Period	Test Period	Accuracy	Precision	Recall	F1-Score	Class 0	Class 1	Class 2
2023-11-26 to 2024-02-26	2024-02-26 to 2024-03-26	0.347	0.330	0.356	0.327	204	156	136
2023-12-26 to 2024-03-26	2024-03-26 to 2024-04-26	0.419	0.388	0.365	0.359	251	144	149
2024-01-26 to 2024-04-26	2024-04-26 to 2024-05-26	0.343	0.314	0.315	0.315	224	121	148
2024-02-26 to 2024-05-26	2024-05-26 to 2024-06-26	0.412	0.394	0.394	0.392	230	167	123
2024-03-26 to 2024-06-26	2024-06-26 to 2024-07-26	0.349	0.424	0.415	0.330	210	114	195
2024-04-26 to 2024-07-26	2024-07-26 to 2024-08-26	0.393	0.386	0.399	0.361	219	96	176
2024-05-26 to 2024-08-26	2024-08-26 to 2024-09-26	0.404	0.356	0.348	0.349	261	103	180
2024-06-26 to 2024-09-26	2024-09-26 to 2024-10-26	0.412	0.341	0.342	0.333	301	142	74
2024-07-26 to 2024-10-26	2024-10-26 to 2024-11-26	0.274	0.317	0.292	0.266	154	210	136
2024-08-26 to 2024-11-26	2024-11-26 to 2024-12-26	0.301	0.376	0.362	0.299	207	102	200

References

[Appel, 2005] Appel, G. (2005). *Technical Analysis: Power Tools for Active Investors*. Financial Times Prentice Hall.

[Basak et al., 2019] Basak, S., Kar, S., Saha, S., Khaidem, L., and Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance*, 47:552–567.

[Breiman, 2001] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

[Fischer and Krauss, 2018] Fischer, T. and Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2):654–669.

[Forex Software, 2025] Forex Software, L. (2025). Historical forex data.

[Glorot et al., 2011] Glorot, X., Bordes, A., and Bengio, Y. (2011). Deep sparse rectifier neural networks. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS 2011)*, pages 315–323.

[Kingma and Ba, 2015] Kingma, D. P. and Ba, J. (2015). Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*.

[Lu, 2024] Lu, Z. (2024). Comparison of stock price prediction models for linear models, random forest and lstm. *Applied and Computational Engineering*, 54:226–233.

[Pan, 2024] Pan, X. (2024). The comparison between random forest and lstm models based on the gold price prediction. *Advances in Economics, Management and Political Sciences*, 94:123–129.

[Riedmiller and Lernen, 2014] Riedmiller, M. and Lernen, A. (2014). Multi layer perceptron. *Machine Learning Lab Special Lecture, University of Freiburg*, 24.

[Sezer et al., 2019] Sezer, O., Gudelek, U., and Ozbayoglu, M. (2019). Financial time series forecasting with deep learning : A systematic literature review: 2005-2019.

[Sparks and Yurova, 2006] Sparks, J. and Yurova, Y. (2006). Comparative performance of arima and arch/garch models on time series of daily equity prices for large companies. *Unive Illinois Chicago*.

[Wilder, 1978] Wilder, J. W. (1978). *New Concepts in Technical Trading Systems*. Trend Research.

[Wu, 2024] Wu, H. (2024). Comparison of random forest and lstm in stock prediction. *Advances in Economics, Management and Political Sciences*, 86:28–34.