INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Effective Federal Funds Rate Forecasting: Deep Learning Application

Ana Rita Ferreira Alão

Master in Data Science

Supervisor:
PhD Vivaldo Manuel Pereira Mendes, Associate Professor,
Iscte - Instituto Universitário de Lisboa

Co-Supervisor:
PhD Diana Elisabeta Aldea Mendes, Associate Professor,
Iscte - Instituto Universitário de Lisboa

September, 2023

Department of Quantitative Methods for Management and Economics/ Department of Information Science and Technology

Effective Federal Funds Rate Forecasting:
Deep Learning Application

Ana Rita Ferreira Alão

Master in Data Science

Supervisor:
PhD Vivaldo Manuel Pereira Mendes, Associate Professor,
Iscte - Instituto Universitário de Lisboa

Co-Supervisor:
PhD Diana Elisabeta Aldea Mendes, Associate Professor,
Iscte - Instituto Universitário de Lisboa

September, 2023

Resumo

Os modelos de *Machine Learning* (ML) podem, de facto, auxiliar a Reserva Federal dos Estados Unidos na formulação da política monetária. Este estudo contribui para a crescente literatura sobre previsões da taxa de juros efetiva dos EUA, fazendo uso de algoritmos de ML, como *Random Forest* (RF), *Light*GBM, *Long Short-Term Memory* (LSTM), *Bilateral* LSTM e *Gated Recurrent Unit* (GRU). Após a aplicação de técnicas de redução de dimensionalidade em uma base de dados com mais de 500 variáveis, foi identificado um conjunto final de variáveis de maior relevância para a previsão, das quais se destacam quatro indicadores financeiros, nomeadamente taxa de juro do papel comercial com *rating* AA a 3 meses, taxa de juro dos bilhetes do tesouro a 3 e 6 meses e o spread entre a taxa de juro do papel comercial a 3 meses e a taxa efetiva do banco central dos EUA. Por sua vez, variáveis relativas ao mercado de trabalho também revelam ter alguma relevância, entre elas o número de trabalhadores na economia americana no setor dos serviços, na industria e o número de trabalhadores no setor privado e agências estatais. Os resultados apontam que, em geral, esses algoritmos apresentam uma *performance* significativamente superior face à simples previsão média da variável dependente e face ao modelo ARIMA. Contudo, vale destacar que apenas a superioridade dos algoritmos de *Deep Learning* (DL) em relação ao ARIMA é estatisticamente significativa, de acordo com o teste de Dielbold-Mariano. Além disso, os algoritmos de DL demonstram superar o desempenho do RF e do LightGBM. A combinação de previsões não apresenta benefícios notáveis para a melhoraria no desempenho dos modelos.

Key-words: Aprendizagem Máquina; Aprendizagem Profunda; Taxa de juro efetiva dos E.U.A.; Previsão; Redes Neuronais

Abstract

Machine Learning (ML) models can indeed assist the Federal Reserve System in conducting monetary policy. This study contributes to the expanding literature on forecasting the effective federal interest rate in the US, utilizing Machine Learning algorithms such as Random Forest (RF), LightGBM, Long Short-Term Memory (LSTM), Bilateral LSTM, and Gated Recurrent Unit (GRU). After applying dimensionality reduction techniques to a dataset comprising over 500 variables, a final set of variables with the most significant relevance for prediction was identified. Among these, four financial indicators stand out: the interest rate on AA-rated commercial paper at 3 months, the interest rates on treasury bills at 3 and 6 months and the spread between the 3-month commercial paper and the Effective Federal Funds (EFFR). Additionally, three labor market indicators were highlighted, including the number of workers in the American economy in the services sector, the industry sector, and the number of workers in the private sector and government agencies. The results indicate that, on the whole, these algorithms significantly outperform both the simple average prediction of the dependent variable and the ARIMA model. According to the Diebold-Mariano test, it is worth noting that only the superiority of Deep Learning (DL) algorithms over ARIMA is statistically significant. Furthermore, DL algorithms overperformed RF and LightGBM. The combination of forecasts does not yield notable benefits in enhancing model performance.

Key-words: Machine Learning; Deep Learning; Effective Federal Funds Rate; Forecast; Neural Networks

Table of Contents

List of tables

List of figures

List of equations

CHAPTER 1

Introduction

The effective federal funds rate (EFFR) is the interest rate at which depository institutions trade federal funds (balances held at Federal Reserve Banks) with each other overnight and it is essentially determined by the market. However, the EFFR is influenced by the Federal Reserve (Fed) through open market operations to reach the federal funds rate target set by the Federal Open Market Committee (FOMC). In other words, Fed implements monetary policy by targeting the EFFR, being of particular interest for various market participants to forecast its value due to the huge impact that monetary policy conduction has in our lives. Furthermore, the ability of market participants to predict the federal funds rate is important to modern analyses of monetary policy, in the sense that it is believed that this rate can act as the benchmark to set other interest rates in the economy through the market's expectation of monetary policy actions that directly affect the federal funds rate.

Indeed, this thematic becomes even more relevant if we consider the current situation that the global economy is facing and, with it, the United States. Inflation levels were not that high for decades, particularly for the US economy, which only reached even higher inflation levels during the 1970's oil crisis (see Figure 1.1.)

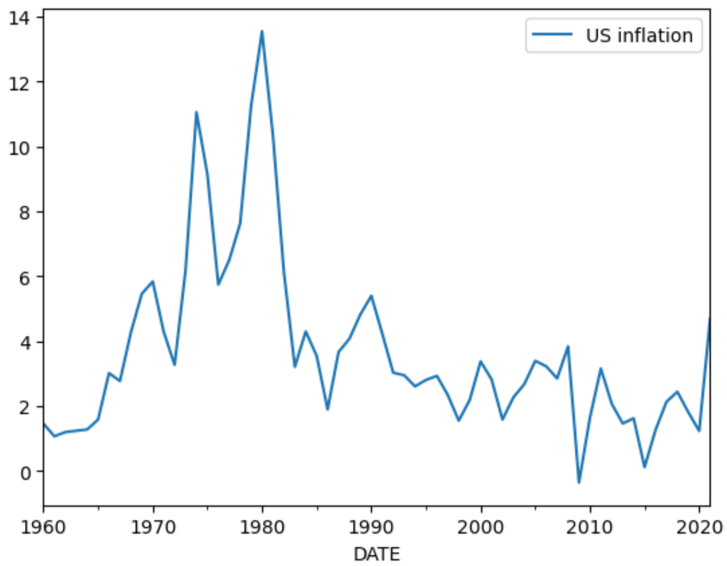

Figure 1.1.: United States inflation (%)

Worldwide, central banks have started to normalize monetary policy to control inflation and bring it back to its target level (International Monetary Fund, 2022). Fed's recent policy is a clear example of this tightening, having raised interest rates several times during 2022, and according to some FOMC members, the job is not done.

Most Open Market Committee members believe that the Fed funds rate will end in 2023 somewhere around 5% (Buckley, 2022). Jerome Powel, Fed's president, warned in November 2022 that it is too soon to start loosening monetary policy and stated that the Fed 'will stay the course until the job is done' (Buckley, 2022).

This uncertainty regarding the evolution of interest rates and inflation itself causes a direct impact on economic agents' decisions, either in investment or consumption decisions. If there is the perspective that interest rates will continue to rally in 2023, families and companies will prefer to save instead of consuming or investing once the price of money is expected to increase. It is then of extreme importance to forecast the fed funds rate path, which corroborates this study's importance.

Concerning the behavior of the federal funds rate, it is necessary to refer to the well-known Taylor rule. This policy rule is a simple function that relates short-term policy rates with inflation and output gap, being the best strategy that policymakers should follow to set and adjust the short-term interest rate in response to some other economic variables (Board of Governors of the Federal Reserve System, 2018). Furthermore, it is a good guide to conduct monetary policy (Leith & Wren-Lewis, 2009), and many authors have been modeling fed funds rate reaction function being aware of this rule, using it as a "baseline model" (see, for example, Hinterlang, 2019; Lee *et al.*, 2015; Hayo & Neuenkirch, 2010 and Monokroussos, 2011).

However, it has been recently proved that the Taylor rule has some issues prescribing short-term interest rates due to its excessive simplicity (Svensson, (2002, 2017, 2020)).

Furthermore, most studies have focused on understanding and modeling monetary policy behavior and forecasting the federal funds target rate set by FOMC. At the same time, very few have been dedicated to studying the effective federal funds rate. It is the latter that reflects market and economic conditions, and it is the one that the Fed intends to manage based on the defined target rate, given its influence on the economy. Thereby, to fill the gap in the literature, this thesis aims to forecast the effective federal funds rate and to incorporate the application of machine learning modeling techniques, particularly understanding the forecasting power of deep learning models.

Moreover, additional study topics, such as dimensionality reduction methods, will be explored to identify the most significant features that can explain changes in the effective federal funds rate, using the available data. Furthermore, in line with the methodologies employed by Hinterlang (2019), Sarno et al. (2005), and Lapp et al. (2003), this report will also conduct Forecast Encompassing Tests among models to assess potential performance improvements resulting from forecast combinations.

Chapter 2 of this thesis provides an overview of relevant concepts for this research. Following that, Chapter 3 presents the state of the art in forecasting the federal funds rate. In Chapter 4, the considered dataset will be analysed and explained, the feature selection strategy employed, the models, hyperparameter tuning, and the evaluation methodology applied. Chapter 5 presents the results, and Chapter 6 concludes the thesis.

CHAPTER 2

Literature Review

Many papers analyze the short-term interest rates target set by central banks, as well as the variables that have the most importance when deciding to increase, maintain, or decrease interest rates (see, for example, Fair, 2005 and Quah & Hemamaline, 2008). At the end of the day, it is the choice of the level of short-term interest rates that will dictate the future of the economy, by affecting overall financial conditions through several channels.

The best-known simple policy rule to express how central banks conduct monetary policy is the Taylor Rule (Taylor, 1993). This rule only takes into consideration inflation and the output gap when defining the nominal short-term interest rate that should be imposed to achieve the inflation target:

$$i_t^* = \pi_t + \theta(\pi_t - \pi_t^*) + \gamma y_t + r^* \tag{1}$$

In the above expressed equation, i_t^* is the target for the short-term nominal interest rate, π_t is the current inflation rate, π^* is the inflation target rate (and so $\pi_t - \pi^*$ is the "inflation gap"). y_t represents the output gap and r^* is the equilibrium level of the real interest rate. θ and γ are positive coefficients representing the weight each central bank attributes to each objective. In the case of the Fed, these two constants will tend to be 0.5, and the equilibrium level of the real interest rate and the inflation target will both be equal to 2 percent, since Fed has price level stability and sustainable economic growth as objectives (Molodtsona *et al.*, 2008).

Despite the recognized ability of this policy rule to determine fed funds target rate movements (Molodtsona *et al.*, 2008), there is a broad consensus among researchers that this rule is too elementary to model such a complex economic reality (Svensson, 2017), thereby, it should only be seen as a mere "guideline" for monetary policy (Svensson, 2002). Furthermore, this reaction function lacks valuable information needed to determine short-term interest rates fixed by central banks, such as forward-looking variables, once it only looks for current information (Svensson, 2017). Because of that, the Taylor rule does not allow for deviations or shocks, since it is considered an incomplete guide (Svensson, 2002). In this context, many authors only use the Taylor rule as a guide to forecast short-term interest rates, constructing new and more sophisticated versions by adding more variables.

Hinterlang, (2019) has recently also analyzed the forecasting performance of monetary policy reaction functions, having the Taylor rule as the baseline model. The author used the US Federal Reserve's Greenbook real-time data, covering a more recent period, 1987:Q3 to 2012:Q4. Hinterlang, (2019) explored both linear and non-linear relationships between the target variable (fed funds rate) and the considered independent variables (lagged fed funds rate, inflation and output gap). Furthermore, the author also included other models commonly used as benchmark ones within the literature, namely the AR and ARIMA (in particular, the author used an AR(2) and an ARIMA(1,1,0)).

For linear relationships, Hinterlang (2019) made some modifications to the original version of the Taylor (1993) rule, including policy inertia along the lines of Clarida *et al.* (1998) with two lagged interest rate terms. By the other side, to study the nonlinearity, the author made use of Smooth Transaction Models (STR), particularly the logistic (LSTR), the logistic-quadratic (L2STR), the exponential (ESTR), and Artificial Neural Networks.

Moreover, the Taylor rule has also served as "baseline model" for other authors. Sarno *et al.* (2005) tried to model the daily behavior of the effective fed funds rate in the last decade of the XX century, and this way, they proposed several models, including a variant of the model proposed by Taylor[1]. By its turn, Kim *et al.*, (2009) presented two versions of a Taylor rule model for setting interest rates that included a well-grounded set of variables with more macroeconomic theory fundamentals. These two versions are distinguished by different measures of inflation: the backward-looking Taylor rule model, which employs last month's core Consumer Price Index (CPI) inflation rate, and the forward-looking Taylor rule model, which utilizes an expected inflation measure (both use employment gap).

However, besides the increasing evidence of the missing explanatory power of this policy rule as discussed before, Sarno *et al.* (2005) and Kim *et al.* (2009)'s results provide evidence that models created as versions or adjustments of the Taylor rule are giving the best forecast performances comparing with other time series regressions.

[1] Under this model, future daily changes in the daily U.S. effective federal funds rate are determine by current deviations of the fed funds rate from its target.

Another topic that has gained more attention among researchers is analyzing which variables are most considered by the Fed when setting the level of short-term interest rates. Furthermore, there is a huge and well substantiated trend among researchers to incorporate the initial set of variables - inflation and output (gap) - in their models to predict the Fed's behavior through time (see, for example, Hauwe *et al.*, 2013; Hu & Phillips, 2004a; Hinterlang, 2019 and Kim *et al.*, 2009).

Although inflation expectations are considered a crucial indicator of monetary policy decisions (Hubert, 2015), surprisingly, inflation (and in some cases, its lagged values) showed not to be an important feature for predicting monetary policymaking (Hauwe *et al.*, 2013; Hu & Phillips, 2004a and Lapp et al., 2003).

Contrary, money growth (M2) has been showed to be a good predictor (Clarida *et al.*, 1998 and Hu & Phillips, 2004a).

Two more variables contain a good predictive power to forecast FOMC decisions. In addition to the management of the interest rate target, FOMC can also produce effects on financial conditions by communicating how it intends to adjust monetary policy shortly. This type of information is called "forward guidance," and it is provided at each FOMC postmeeting communication. Therefore, the public tends to create expectations about FOMC monetary policy decisions. Indeed, FOMC postmeeting statements are shown to produce good predictions of how FOMC will change interest rates (Hayo & Neuenkirch, 2010; Pakko, 2005 and El-Shagi & Jung, 2015).

Moreover, there is also some evidence that financial indicators provide as good or better predictions as the Taylor rule, such as the 6-month Treasury bill spread relative to the federal funds rate and two-year Bond yields (Piazzesi, 2005; Kauppi, 2012 and Lapp, 2003).

There is a huge trend to apply a recursive one-step-ahead strategy to forecast the setting of interest rates (Kim *et al.*, 2009; Hinterlang 2019; Sarno *et al.* 2005; Hauwe *et al.* 2013 and Seibert *el al.*, 2021). This forecast strategy[2] is the most intuitive strategy for multi-step-ahead time series forecasting. However, the recursive strategy may produce unsatisfactory results when compared with other forecasting strategies, such as direct or rolling windows. The major concern regarding this wide-preferred and used technique results from the introduction of error each time a predicted observation is added to the

[2] This a forecasting approach where we make predictions for each future time step by using the previously predicted values. In simpler terms, it's like taking one step at a time (Yan-jie et al., 2017).

sample to predict the next observation. This means that it will accumulate errors as the forecast horizon increases (Yan-jie *et al.*, 2017).

Another technique also showed to be a trend among researchers was the test for equal predictive accuracy by Diebold & Mariano (1995) to investigate the statistical significance of the forecasting results, as employed, for example, by Hinterlang (2019) and Richardson et al. (2021). Furthermore, the forecasting combination also appears to be a choice among authors. However, this technique did not improve performance results drastically (Hinterlang, 2019 and Sarno et al., 2005).

Lastly, applying ANN to forecast the Fed's behavior regarding the setting of interest rates is not a very explored field among researchers, meaning it is still an unexplored study subject. In this context, Hinterlang (2019) proposed a *single-hidden-layer* recurrent neural network with a hyperbolic tangent activation function. The choice of the number of neurons in the hidden layer was done by phases. First, after dividing the TS into training and validation sets (80% - 20%, respectively), Hinterlang (2019) estimated[3] the ANN looping over 1 to 10 hidden units with 30 different randomly chosen initial weights and biases each. Then, the number of hidden units that minimized the resulting mean squared error in the validation set averaged over the 30 trials was chosen to be the optimal one. Lastly, by having fixed this optimal number, the trial with the lowest validation mean squared error served as the optimal initial weights and biases. As a result, the chosen numbers of hidden units were 4, 1 and 2 for the within-quarter, the backward-looking and the forward-looking version, respectively.

Quah & Hemamaline (2008) focused on the prediction of the direction and magnitude of changes in Fed funds target rate by exploring two-level neural network architecture. The first level was the Self-Organizing Map (SOM), which is used almost like a "sample selection" processor because it will classify the data into those that impact the Fed funds target rate and those that do not. The second level was the General Regression Neural Network (GRNN), which is going to be used to forecast the direction and magnitude of fed funds target rate changes based on the data presented to it.

Quah & Hemamaline (2008) used the *single-hidden-layer* and chose a liner activation function. One of their main conclusions was that "sample selection" done by the SOM network aids in better predicting the Fed funds target rate changes.

[3] Using Levenberg-Marquardt algorithm (LMA).

CHAPTER 3

Background

3.1. Machine Learning

Since early, some studies have been focusing on Machine Learning (ML), particularly Samuel (1959), which has concentrated on developing learning procedures applied to games. Time after, Mitchell (1997) contributed to the theory that has formed the core of machine learning and the theoretical understanding of learning by presenting some key learning algorithms (e.g., Decision Tree Learning, Artificial Neural Network (ANN), and Bayesian Learning).

Machine Learning (ML) is a branch of computer science that broadly aims to enable computers to "learn" without being directly programmed (Samuel, 1959). This concept of ML started with ideas of developing algorithms that can learn and gain some knowledge from the (past) experience to predict future behaviors, being able to continuously improve their learning behavior (Holzinger, 2017). In its turn, this learning concept can be viewed as the task of looking within a huge set of possible explanatory variables (features) and finding the ones that best explain/fit the data (Mitchell, 1997, p. 23).

In ML there are four commonly used learning methods, each of which aims to solve different problems: *supervised*, *unsupervised*, *semisupervised*, and *reinforcement learning*. Focusing on the first two, in a brief way: *Supervised Learning* is when we are predicting an outcome value (e.g., the dependent variable or the variable of interest), often called its "label", which value is known *a priori* for each observation in the dataset (e.g., regression models). In opposition, in *Unsupervised Learning*, the label outcome is unknown; the algorithm tries to identify patterns, relationships, and groups within the input data without knowing the true outcome. For example, clustering and dimensionality reduction are among the most common uses of unsupervised learning algorithms (Choi *et al.*, 2020; Bi *et al.*, 2019; Masini *et al.*, 2021).

In *Supervised Learning,* there is a vast range of different methods to model data. The choice between them is considered one of the issues of Machine Learning (Mitchell, 1997, p.15). It must consider the relationship between the dependent and explanatory variables (e.g., linear or non-linear); the problem type - regression or classification, and the most adequate algorithm.

Some common ML algorithms are artificial neural networks, support vector machines, decision trees, k-means clustering and naïve Bayes, as described by Bi *et al.*, 2019.

3.2. Deep Learning

Deep Learning (DL), an ML technique to deal with more complex patterns and problems, has been drawing some attention from many researchers (Maleki *et al.*, 2020). This has heavily contributed to the increasing popularity of ML in a broader community (Holzinger, 2017). Artificial Neural Networks (ANN) are one of the traditional non-linear algorithms behind deep learning (Masini *et al.*, 2021) and are inspired by biological neural connections (Choi *et al.*, 2020).

ANN contains a layer of input neurons, where each neuron corresponds to an input variable. This is the raw information that the neural network receives, which aims to understand and learn from data to predict the desired variable. From the input layer, each feature gets a weight according to the degree to which it influences the target. Then, it will connect with each neuron of the following (hidden) layer according to the weight attributed. For instance, in the consecutive hidden layers, each node (neuron) will transform the received weighted information into a desired output, by applying an activation function (Bi *et al.*, 2019). Thus, each neuron in the hidden layer is a computational unit that first calculates a linear function of its inputs and weights and then applies a non-linear or a linear transformation (the activation function) to the resulting value (Maleki *et al.*, 2020 and Choi *et al.*, 2020). Figure 3.1., illustrates a 3-layer neural network, where the units are represented by circles and the weights by arrows.

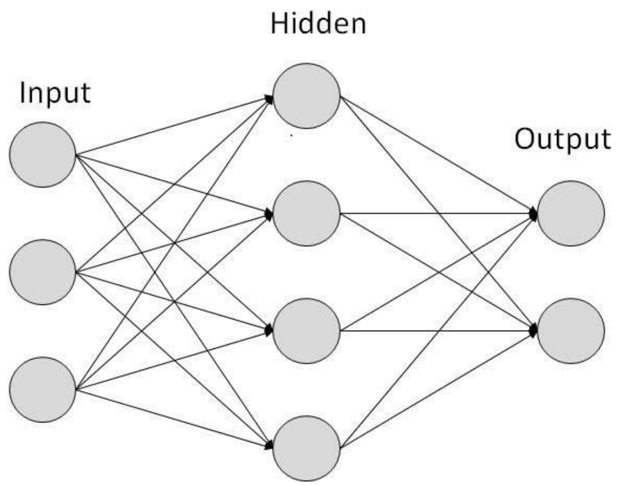

Figure 3.1.: A simple neural network architecture

The activation function could be a sigmoid, hyperbolic tangent, identity, or a rectified linear unit (Choi *et al.*, 2020). What is essential is to introduce nonlinearity into the training process, otherwise, even a complex neural network with multiple neurons would be unable to learn and reproduce non-linear patterns and relationships (Maleki *et al.*, 2020).

The initial weights are set randomly, and the provided network output is compared with the target values using a loss function, which is merely the difference between predicted and true values. The goal is to minimize the loss function to zero to predict the outcome values as accurately as possible, being aware of the overfitting tendency problem (Bi *et al.*, 2019).

In response to the loss function value, there could be a 2-fold optimization process, where the resulting error values are distributed backward through the network to assign an error value contribution to each layer (called "back-propagation"). At the same time, there is an adjustment/update of the weights to minimize the loss function ("weight adjustment") (Bi *et al.*, 2019).

Although the advantage of ANN in learning non-linear and complex problems, Bi *et al.*, (2019) pointed out some disadvantages of using these algorithms, such as the fact that their performance is high dependent on their architecture (which includes the number of hidden layers and neurons, the activation function chosen, among others) and on their training parameters (such as the stopping criteria applied), what makes it harder to achieve a good performance given the several possible combination of parameters. Moreover, due to this complexity, technical expertise is needed to set and tune these parameters. Another disadvantage of ANN, according to Bi *et al.*, (2019), concerns the difficulty of interpreting the individual variable effects on the target since they suffer from some transformations.

3.3. Models

3.3.1. Random Forest

The Random Forest (RF) algorithm, whether for regression or classification problems, was initially developed by Ho (1995) and Breiman (2001). It involves a combination of multiple predictor trees at training time, each based on independent sub-samples (if Bootstrap == True) of the original dataset.

This approach helps reduce the mean squared forecast error compared to using a single individual tree to fit the data, as it benefits from averaging across several independent weak learners/predictor trees (in regression tasks). In classification problems, the output class is the mode of the classes. Thus, this algorithm reduces variance and makes it less suitable for overfitting (which is the main motivation behind RF).

3.3.2. LightGBM

LightGBM is a gradient boosting ensemble method that, as RF algorithm, combines weak learners (decision trees) to create a stronger predictive model. But, unlike RF, the final prediction value is not the mean/mode from the trained whole trees. Boosting algorithms iteratively build a series of models, each focusing on correcting the errors of the previous ones. Therefore, it starts training with only a single tree and then trains other model/tree with the errors from the previous one, and so on.

3.3.3. Long short-term memory

Long Short-Term Memory (LSTM), introduced by Hochreiter & Schmidhuber (1997), is a type of Recurrent Neural Network (RNN) that is specifically designed to handle sequential data, such as time series, and to learn long-term dependencies. The key distinction between vanilla RNNs and LSTMs is that the latter supports the gating of the hidden state that controls the flow of information within the network. These gates enable LSTMs to selectively update and retain information, making them more effective at capturing relevant patterns while ignoring irrelevant information. These gates include the input, forget, and output gates.

Regarding the Forget gate (see Figure 3.2.), given the previous hidden state (H_{t-1}) and the current information (X_t), it is going to compute weights from 0 to 1 with a sigmoid function for each number in the previous cell state (C_{t-1}) to decide what information is going to be thrown away from the current cell state (C_t). Forget gate is defined as follows:

$$F_t = \sigma(W_F \cdot [H_{t-1}, X_t] + b_F) \quad (2)$$

where W_F is the weight parameter, b_F is the bias and the symbol $\odot$ is the Hadamard (elementwise) product operator.

Next, a tanh activation function creates a vector of new candidate values, $\tilde{C}_t$, where the Input gate layer is going to act by also applying a sigmoid function to attribute weights from 0 to 1 for each value of this vector to decide what new information is going to be stored in the current cell state (C_t). Both candidate values and Input gate are defined as follows (respectively):

$$\tilde{C}_t = \tanh(W_C \cdot [H_{t-1}, X_t] + b_C) \tag{3}$$

$$I_t = \sigma(W_I \cdot [H_{t-1}, X_t] + b_I) \tag{4}$$

where W_I and W_C are the weights parameters and b_I and b_C are the bias. Therefore, after applying the weights from the Forget and Input gates, LSTM has a new cell state, C_t:

$$C_t = F_t \cdot C_{t-1} + I_t \cdot \tilde{C}_t \tag{5}$$

Suppose that the forget gate is always 1 and the input gate is always 0, the memory cell's internal state (C_{t-1}) will remain constant forever, passing unchanged to each subsequent time step. However, input and forget gates allow the model to learn when to keep this value unchanged and when to perturb it in response to subsequent inputs.

Lastly, looking at the last part composing Figure 3.2., the Output gate layer is going to compute the output of the memory cell, by creating a filtered version of the current memory (C_t). As the other gates, the Output gate is a sigmoid function to attribute weights from 0 to 1, defined as follows:

$$O_t = \sigma(W_O \cdot [H_{t-1}, X_t] + b_O) \tag{6}$$

where W_O is the weight parameter and b_O is the bias.

So, a sigmoid layer is applied to decide what parts of the cell state it is going to output. It forces the cell state values to be between -1 and 1 trough tanh and is multiplied by the output of the sigmoid function so that LSTM only outputs the parts it decided to. The output hidden state is defined as follows:

$$H_t = O_t \cdot \tanh(C_t) \tag{7}$$

Whenever the output gate, O_t, is close to 1, it allows the memory cell's internal state to impact the subsequent layers uninhibited. In contrast, for output gate values close to 0, LSTM prevents the current memory from affecting other network layers at the current time step.

Note that a memory cell can accrue information across many time steps without impacting the rest of the network (as long as the output gate takes values close to 0), and then suddenly impact the network at a subsequent time step as soon as the output gate flips from values close to 0 to values close to 1.

As illustrated in Figure 3.2., the memory cell output (H_t) will be used as input for the next cell.

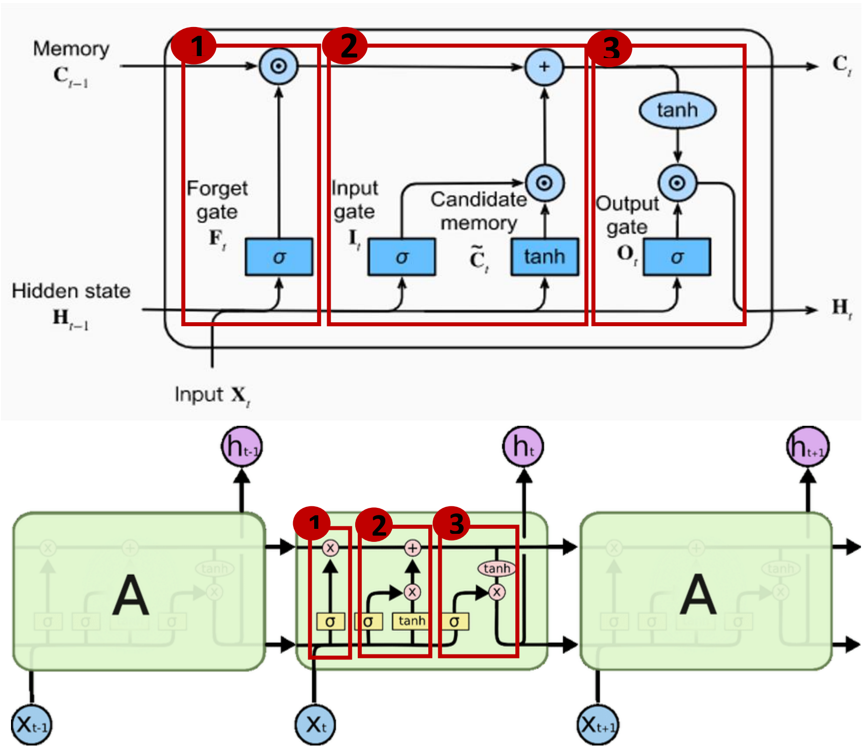

Figure 3.2.: LSTM cell architecture

3.3.4. Bidirectional long-short term memory

A Bidirectional LSTM (Bi-LSTM) is a variant of the LSTM neural network that processes input data in two directions: from the beginning to the end (forward direction) and from the end to the beginning (backward direction). Therefore, one of the advantages of Bi-LSTMs is the effective increase in the amount of information available to the network learn. They can be trained using all available input information from the past and the future within a specific time frame.

3.3.5. GRU

A Gated Recurrent Unit (GRU), introduced by Cho et al. (2014), is a type of recurrent neural network (RNN) architecture, similar to LSTM, designed to capture long-range dependencies in sequential data while mitigating some of the complexities associated with LSTMs (e.g., the increased computational demands).

In GRU, the LSTM's three gates are replaced by the Reset and Update gates. The first one controls how much of the previous state H_{t-1} (past information from the previous time steps) the model might still want to remember. In this way, the model becomes capable of eliminating the risk of vanishing gradient problem by having the ability to copy all the information from the past. Likewise, the Update gate would allow it to control how much of the new state is just a copy of the old one. Both gates receive as input the current time step (X_t) and the hidden state of the previous time step (H_{t-1}) and a sigmoid activation function is applied to squash the result between 0 and 1, as in LSTM's gates. Therefore, the Reset (R_t) and the Updated gate (U_t) are defined as:

$$R_t = \sigma(W_R \cdot [H_{t-1}, X_t] + b_R) \tag{8}$$

$$U_t = \sigma(W_U \cdot [H_{t-1}, X_t] + b_U) \tag{9}$$

where W_R and W_U are weight parameters and b_R and b_U are bias parameters.

After, by applying the reset gate weights to the past information (H_{t-1}), it will determine what to remove from the previous time steps. It uses a tanh activation function to get a candidate hidden state, $\widetilde{H}_t$:

$$\widetilde{H}_t = \tanh(W_H \cdot [(R_t \cdot H_{t-1}), X_t] + b_H) \tag{10}$$

where W_H is a weight parameter, b_H is the bias. Note that the result is a candidate since it still needs to be incorporated into the action of the update gate (U_t).

Finally, to get the new hidden state (H_t) it is crucial to know what is the degree to which this new hidden state aligns with the previous state, H_{t-1}, as opposed to its resemblance to the candidate state, $\widetilde{H}_t$. This is determined by the Update gate, U_t:

$$H_t = U_t \cdot H_{t-1} + (1 - U_t) \cdot \widetilde{H}_t \tag{11}$$

Note that whenever the update gate U_t is close to 1, the algorithm retains the old state H_{t-1}. In this case, the current information X_t is completely ignored, effectively skipping time step t in the dependency chain. By contrast, whenever U_t is close to 0, the new latent state H_t approaches the candidate's latent state $\widetilde{H}_t$.

Figure 3.3. illustrates the GRU cell architecture explained above.

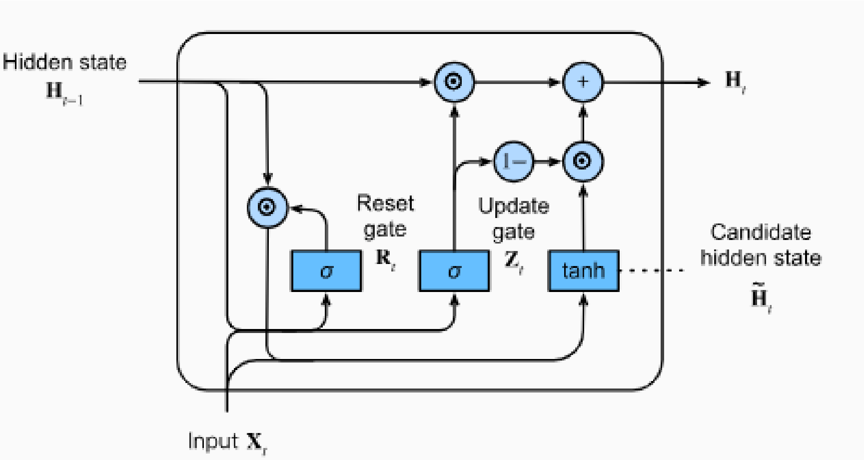

Figure 3.3.: GRU cell architecture

3.4. Monetary Policy

The Federal Reserve conducts monetary policy by managing the level of short-term interest rates and, thereby influencing the availability and cost of money and credit in the economy. Fed, particularly, does that to achieve three main goals: maximum employment, stable prices and moderate long-term interest rates in the US (Board of Governors of the Federal Reserve System, 2016)

The federal funds rate is the interest rate at which depository institutions trade federal funds with each other overnight. When a depository institution has surplus balances in its reserve account, it lends to other banks needing larger balances, because, by law, banks must maintain a reserve equal to a certain percentage of their deposits in an account at a Federal Reserve bank. This rate that the borrowing institution pays to the lending institution is called the effective federal funds rate and is the weighted average rate for all these types of negotiations.

According to FOMC's decision regarding the federal funds target rate, which depends on what FOMC believes will be the state of the economy, the Fed would influence the effective federal funds rate through open market operations or by buying and selling government bonds (government debt). For instance, if the FOMC believes the economy is growing too fast and inflation pressures are inconsistent with the dual mandate of the Federal Reserve, the Committee may set a higher federal funds rate target by selling government bonds to temper economic activity. In the opposing scenario, the FOMC may set a lower federal funds rate target by buying government bonds to spur greater economic activity.

Therefore, the EFFR is crucial for US capital market participants as it indirectly affects other interest rates throughout the economy. More simply, it is the EFFR that the Fed influences to produce changes in the longer-term interest rates, stock prices and the foreign exchange value of the dollar, which then affects a wide range of spending decisions made by households and businesses. Since FOMC's decisions directly impact society´s well-fare, it is of utmost importance to understand what motivates the Federal Open Market Committee to take such decisions.

There are two aspects that are important to have in mind. The first one is that people's expectations play a crucial role when it concerns central bank decisions regarding interest rates. For example, if Federal Reserve communications lead the public to think that the FOMC will soon reduce the federal funds rate to a lower level than previously expected, there are the same consequences that would happen if the FOMC actually reduced its target for the federal funds rate. Indeed, even if the FOMC does not intent to change the target rate at all, if there is only an expectation of it in the market, this will lead to the same effects on the short-term interest rate as if there was an actual change, in what is called the "self-fulfilling expectations". So, in addition to adjusting the target for the federal funds rate, FOMC can also impact financial conditions through its postmeeting statements (where it communicates how FOMC intends to adjust policy in the future) by influencing people's expectations.

The other aspect is that monetary policy actions influence inflation and employment with a lag, which means that an FOMC policy decision will not affect the economy immediately. By that, policymaker's decisions are not only based on the evaluation of the current economic conditions, but also on the forecast of the state of the economy over the next few years. This is why it is believed that monetary policy is forward-looking.

CHAPTER 4

Empirical Application

This chapter will provide detailed information about the database, its utilization for training, validation, and testing, and the variable selection process. Additionally, it will describe the time series and machine learning models used for predicting the effective interest rate of the United States central bank, along with the methodology for selecting key hyperparameters and evaluating the forecast.

4.1. The data

The data consists of a large macroeconomic database (FRED-MD) of 134 monthly US indicators updated monthly using the FRED database[4]. The data ranges from January 1960 to March 2023, resulting in 759 observations.

In addition to these variables, eight other variables were added from the FRED (Federal Reserve Economic Data) database, considered theoretically relevant to explain the effective interest rate of the Federal Reserve (FED). These additional variables are described in Appendix A, Table A1.

For this study, the available data will be divided into train, validation, and test sets, as illustrated in Figure 4.1., with the last 76 observations considered the testing period.

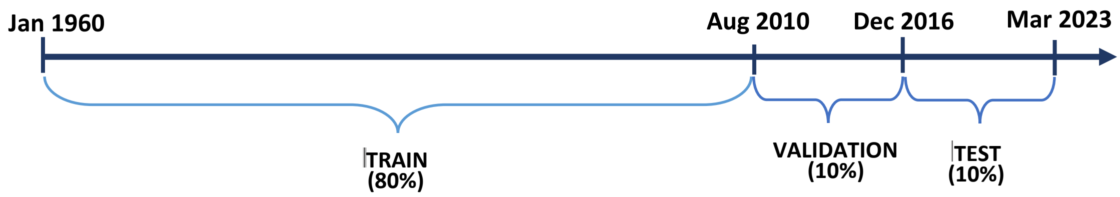

Figure 4.1.: Data division into training, validation and test sets

4.2. Feature Selection

The high dimensionality of the data is a significant concern in this work, mainly when dealing with over 130 variables across more than 750 observations. Using a model with such many variables can lead to overfitting, hindering its ability to generalize to unseen data.

[4] https://research.stlouisfed.org/econ/mccracken/fred-databases/

Additionally, selecting the final set of variables plays a crucial role in the performance of the models. To address this, various feature selection techniques are used. The goal is not only to retain a set of features that individually explain the dependent variable the most effectively but also to identify the set of features that, when combined, achieve this objective.

Consider all the techniques and algorithms applied for feature selection, namely Recursive Feature Elimination (RFE), Interaction Weight based Feature Selection (IWFS) (Zeng et al. 2015), Fast Correlation-Based Filter (FCBF) (Yu and Liu 2003), Minimum Redundancy Maximum Relevance (MRMR) (Peng et al. 2005), Sequential Feature Selection Backwards and Forwards (SFSF and SFSB), Impurity decreases from a Random Forest (RF) and Variance Inflation Factor (VIF).

The first filtering was done by retaining the exogenous features that were selected more than once by several methods (see Appendix A, Table A2.), resulting in the following set of variables: 3-Month AA Financial Commercial Paper Rate (*CP3Mx*), 3 and 6-Month Treasury Bill (*TB3MS* and *TB6MS*), 3-Month Commercial Paper Minus FEDFUNDS (*COMPAPFFx*), S&P's Composite Common Stock: Price-Earnings Ratio (*S&P PE ratio*), Inflation Expectation (*MICH*) and Brave-Butters-Kelley Real Gross Domestic Product (*BBKMGDP*).

Additionally, Table 4.1., reports the Root Mean Squared Error (RMSR) and the Mean Squared Error (MSE) after applying a simple LSTM model (and with no hyperparameters optimization)[5] over the testing period and using only the top 5 features chosen by each of these techniques. The goal is to find the method(s) that produce better forecast results to determine the most compelling feature subset for the time series model.

Table 4.1.: *RMSE and MSE using the top 5 variables from each feature selection technique*

Metrics	RFE	IWFS	FCBF	MRMR	SFSF	SFSB	RF	VIF
RMSE	**0.1910**	0.3873	0.2184	0.4061	0.7053	0.8805	0.3364	3.5189
MSE	**0.0365**	0.1500	0.0477	0.1649	0.4975	0.7753	0.1132	12.3828

According to the results reported in Table 4.1., the method that yields better results is RFE. Therefore, the five most important variables chosen by the Recursive Feature Elimination method will be included in the previous set, which consists of three new

[5] 64 neurons and 1 hidden layer, 100 epochs and a batch size equal to 32.

features, Goods-Producing Industries (*USGOOD*), Total nonfarm (*PAYEMS*), and Service-Providing Industries (*SRVPRD*).

According to this final set of variables, described in Appendix A, Table A4 and Figure A2, financial indicators, namely interest rates, play a crucial role in predicting the EFFR, accounting for 40% of the selected features in this set. This finding is consistent with previous studies by Piazzesi (2005), Kauppi (2012), and Lapp (2003), which also demonstrated that Federal Reserve decisions on how to manage the EFFR respond to changes in bond yields and Treasury bills. For instance, when there is heightened credit risk or uncertainty in the financial markets, the spread between commercial paper rates and the EFFR widens, reflecting the increased cost of borrowing for corporations. This can reduce the amount of money available in the financial system and, in turn, influence the EFFR, causing it to rise. On the other hand, when the yield on treasury bills rises significantly above the EFFR, it creates arbitrage opportunities for financial institutions. They can borrow funds at lower EFFR rates and invest in higher-yielding T-Bills, aligning the two rates more closely.

Additionally, labor market indicators, namely *USGOOD*, *PAYEMS*, and *SRVPRD*, also carry a significant weight (30%) in this final feature set, indicating their predictive power over the EFFR. In particular, the number of US workers in the economy, excluding proprietors, private household employees, unpaid volunteers, farm employees, and the unincorporated self-employed (*PAYMES*) accounts for approximately 80% of the workers who contribute to the Gross Domestic Product (GDP). Therefore, this metric offers valuable insights into the present economic landscape as it reflects the net change in employment within an economy. An upswing in employment figures may signify that businesses are actively recruiting, implying potential business expansion and a more robust economy. Furthermore, individuals newly entering the workforce experience a boost in their personal incomes, assuming all other factors remain unchanged, leading to increased disposable incomes. This, in turn, fosters additional prospects for economic growth (US Bureau of Labor Statistics, 2023). When the economy expands, and unemployment decreases, it can lead to inflationary pressures. Central banks, including the Federal Reserve in the US, may respond to these pressures by increasing interest rates, including the federal funds rate, to curb inflation. This is because higher employment levels can lead to increased consumer spending and aggregate demand. However, the health of the labor market and overall economic conditions not only impacts the EFFR by affecting the Federal Reserve's policy decisions, but also by affecting market expectations and interest rate dynamics.

Stock Market, represented by the *S&P PE ratio* variable, can influence investor behavior and market sentiment, indirectly affecting the central bank's monetary policy decisions, thereby impacting the federal funds rate.

Inflation expectations (*MICH*) gathered by the University of Michigan in its Consumer Survey are among the most important variables to explain EFFR changes, as stated by other authors, such as Hayo & Neuenkirch (2010). This is not surprising because Central Banks, particularly the Federal Reserve, often have an inflation target as part of their mandate. Their expectations matter because they can affect the actual inflation, as explained in section 2.4., the "*self-fulfilling expectations*". For instance, if people expect prices to rise, they will be more willing to buy and invest now rather than save. This increased spending leads to higher demand, which, in turn, tends to drive up the general price level. This would happen even if their expectations do not actually materialize. Therefore, to maintain its inflation target (2%), the Fed would need to change the federal funds rate.

Lastly, the *BBKMGDP*, which represents the monthly GDP growth for the US, can influence the federal funds rate through its impact on monetary policy decisions. GDP growth is closely tied to a country's overall economic activity level, making it a crucial factor in maintaining both price stability and full employment, which fall under the Central Bank's Dual Mandate.

4.2.1. Granger Causality test

The second filtering concerns applying the Granger Causality Test[6] to test if the past values of these TS retained have a statistically significant impact on the current and future values of the target series. To do so, it is necessary first to ensure that the time series are stationary, and therefore, it was applied the Augmented Dickey-Fuller test, whose not rejection of the null hypothesis suggests the presence of a unit root in the time series sample. The results indicate that *TB3MS*, *TB6MS*, *USGOOD*, *PAYMEMS*, *SRVPRD*, and *FEDFUNDS* are nonstationary time series, assuming a significant level of 5%. Regarding

[6] The Null for Granger causality test assumes that all coefficients of the lagged terms of variable X in the equation of Y are null. So, we test if X does not Granger cause Y. If we reject the null, then, we conclude that X Granger causes Y, which means that, variable X has impact on the forecast of variable Y.

this, the logarithmic and the first difference transformations were applied to the mentioned time series, to make them stationary and be able to test Granger causality.

According to the results (see Appendix A, Table A4.), and assuming either a significance level of 5% or 10%, the last five quarters (15 lags) of all time series appear to Granger causes the EFFR. In other words, past information regarding these variables helps to predict the effective federal funds rate, which means that none will be removed from the feature set at this stage. However, with 1%, S&P's Composite Common Stock: Price-Earnings Ratio does not Granger causes the EFFR, not being helpful to predict the effective federal fund rate.

4.2.2. Feature's Correlation

The last round of filtering was based on the correlation between variables, resulting in removing those that, due to strong correlations with others, contribute little or add no useful information for the algorithm to learn. Otherwise, the model might suffer from multicollinearity by having high correlations between independent variables. The linear correlation coefficient matrix is presented in Appendix A, Fig A5.

Following the Chan YH (2003) guideline on the strength of the linear relationship, the target variable has a strong correlation with the financial indicators (*CP3Mx*, *TB3Ms* and *TB6Ms)* and with two of the three labor market indicators, namely *SRVPRD* and *PAYEMS*, since their correlation coefficient is greater than 0.8. These strong correlations are also statistically significant at a 1% level. However, these five time series strongly correlated with the target variable are also strongly correlated with each other, meaning that we will keep one of them for modeling purposes since they represent the same information for understanding the target behavior. Therefore, *CP3Mx* was the only one kept.

Moreover, it's worth noting that the variable *BBKMGDP* exhibits the lowest statistically significant correlation coefficient (0.155), indicating a weak correlation with the target variable, as per Chan YH's (2003) guideline.

The selection of variables to be included in the modeling phase will be made manually from the remaining set[7], based on the combination of variables and their lags that produce the most competitive results in terms of prediction, analysed on the validation set for each model.

[7] *CP3Mx*, *COMPAPFFx*, *S&P PE ratio*, *MICH*, *BBKMGDP*, *USGOOD*.

The following flowchart, presented in Figure 4.2, describes the entire process of filtering variables with potential for inclusion in the models.

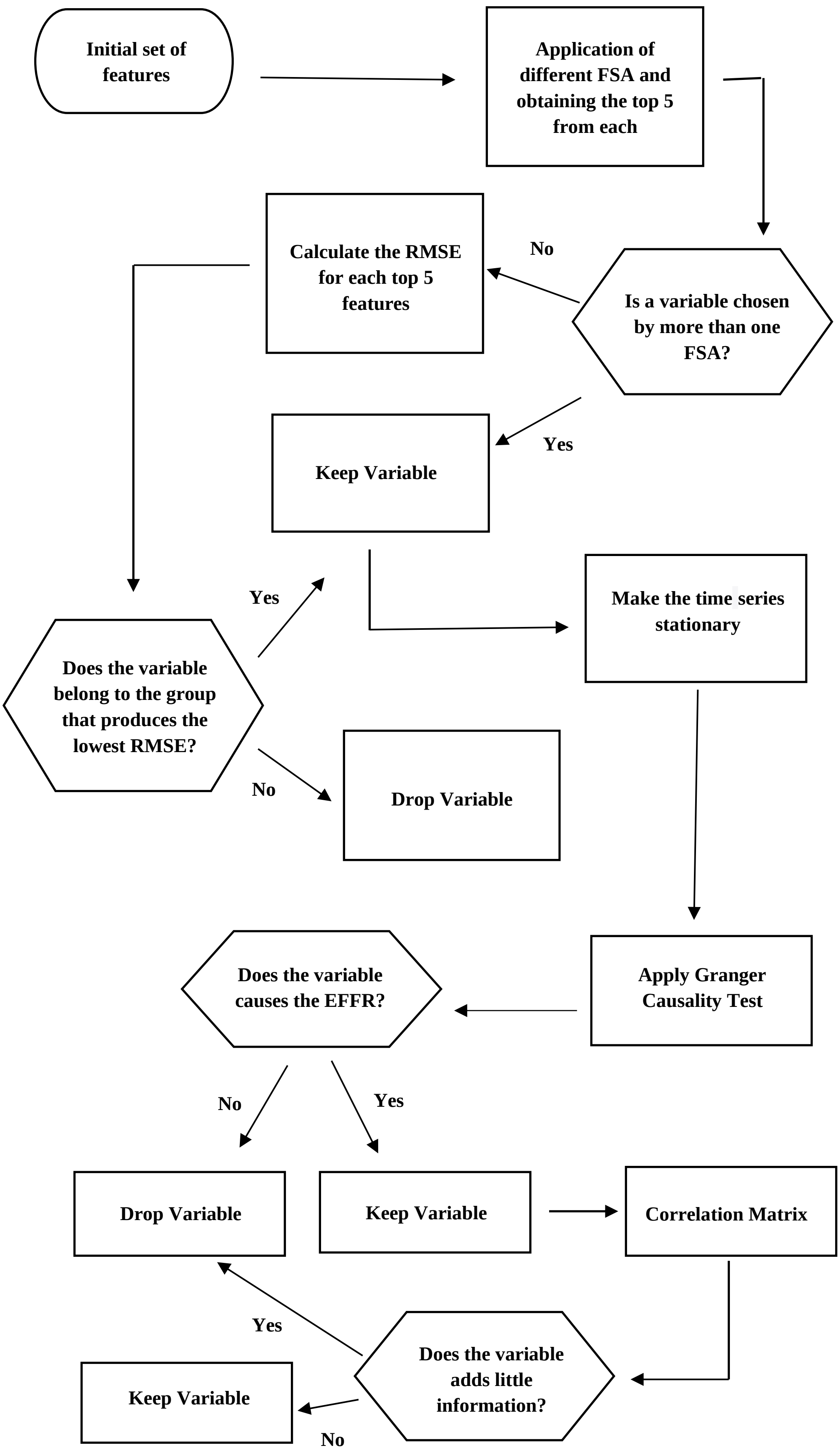

Figure 4.2.: Feature Selection process

4.3. Modeling

Once again, to systemize the employed methodology, all the steps followed in the modeling phase are described in the flowchart illustrated in Figure 4.3.

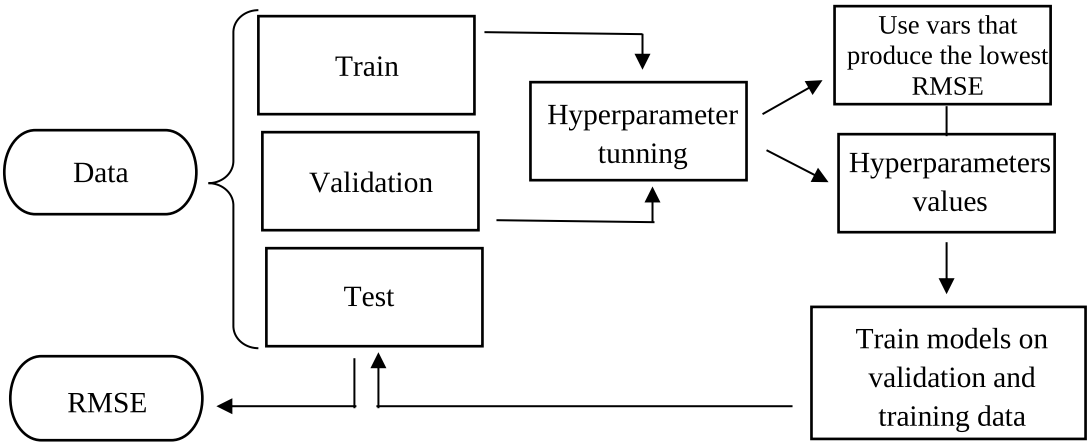

Figure 4.3.: Modeling Process

Below, we briefly describe the most important steps – that conduct a successful modeling procedure.

4.3.1. Hyperparameter tuning and model selection

To determine the optimal hyperparameters, as illustrated in Figure 4.3, we divided the data into training (80%), validation (10%), and test (10%) sets. The selected hyperparameters were based on the best performance achieved during validation.

The optimization process is performed 23 times[8], for each chosen machine learning model, varying the subset of variables and their lags included in the model. Optimal hyperparameter selection was done using the OPTUNA[9] library to minimize the mean squared error (MSE) on the validation set. For ARIMA model, the optimal hyperparameter search was conducted using the pmdarima[10] library. The test set is reserved for evaluating the best model's performance, as the hyperparameters are selected based on results from the validation set.

---

[8] A different model per each independent variable with 1 lag and the dependent variable with also 1 lag, which sums up to 10 models. The same procedure is carried out but considering 2 lags, leading to a total of 20 models. The other 3 models only uses the target variable with 1, 2 and 3 lags.

[9] https://optuna.org/

[10] https://pypi.org/project/pmdarima/

Table 4.3., presents the hyperparameters chosen to optimize for each model, the exploration range, the selected values, and the variables included in the models that yielded these final values.

Table 4.2.: *Variables, hyperparameters ranges and optimal values*

Model	Variables	Hyperparameter	Range	Optimised Value
ARIMA	Fedfunds	p	-	2
		d	-	1
		q	-	1
RF	Fedfunds lag1	No. of trees	2 to 20	2
	Fedfunds lag2	Max depth tree	2 to 8	6
		Min samples split	2 to 10	10
		Min samples leaf	1 to 4	2
		Max features	Auto/sqrt/log2	Auto
LightGBM	Fedfunds lag1	Boosting Type	Gbdt/dart/goss	Goss
	CP3Mx lag1	No. of trees	5 to 50	20
		Max depth tree	2 to 8	5
		Learning rate	0.01 to 0.1	0.0997
		Subsample	0.5 to 1	0.8383
		No. of leaves	5 to 20	19
		Col sample by tree	0.5 to 1	0.8912
LSTM	Fedfunds lag1	Optimizer	Adam/Nadam/RMSprop/SGD	RMSprop
	BBKMGDP lag1	No. of units	10/20/30/40/50	20
		No. of epochs	10/20/30/40/50	30
		Batch size	2/4/8	2
		Dropout rate	0.1 to 0.6	0.1270
		Learning rate	0.001 to 0.8	0.0012
Bi-LSTM	Fedfunds lag1	Optimizer	Adam/Nadam/RMSprop/SGD	RMSprop
	Fedfunds lag2	No. of units	10/20/30/40/50	30
		No. of epochs	10/20/30/40/50	30
		Batch size	2/4/8	2
		Dropout rate	0.2 to 0.6	0.3598
		Learning rate	0.001 to 0.8	0.0018
GRU	Fedfunds lag1	Optimizer	Adam/Nadam/RMSprop/SGD	Nadam
		No. of units	10/20/30/40/50	40
		No. of epochs	10/20/30/40/50	20
		Batch size	2/4/8	4
		Dropout rate	0.3 to 0.6	0.3200
		Learning rate	0.001 to 0.8	0.0010

To prevent overfitting the training data, the hyperparameter search was constrained to a range of ‘small and finite’ values, avoiding the creation of overly specific and complex models. Additionally, a dropout rate was considered for neural networks to mitigate overfitting. The dropout rate determines the probability of temporarily deactivating a node and all its associated forward and backward connections, effectively creating a modified network architecture from the original one. This measure helps prevent the model from learning potential statistical noise and enables better generalization on unseen data. In line with this, all Recurrent Neural Networks models employed in this dissertation have only one hidden layer. Being aware of its importance, note that both LSTM and GRU algorithms assign the highest priority to this hyperparameter during the optimization (see Appendix A, Figures A3 and A4).

4.3.2. Forecasting evaluation methodology

The ARIMA model has an advantage over the others. We are training the ARIMA on data that would have been available to the forecaster in real time and using it to assess the real-time prediction that the forecaster would have made. The other models do multi-step ahead forecasts only with the information presented in train and validation data.

To compare the model's predictive performance, we compute the root mean squared error (RMSE), defined as:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}(y_t - \hat{y_t})^2}{T}} \tag{12}$$

where y_t and $\hat{y_t}$, are the actual and the forecasted values of the effective federal funds rate, and T is the total number of forecasts, which in this case is 76 (months).

The forecasts of a univariate model (in this case, ARIMA) provide the primary benchmark for comparisons between models. The Diebold–Mariano (Diebold & Mariano, 1995) test was used to determine whether the forecasts obtained from each ML model differed significantly from those from the ARIMA model.

Assuming a significance level of 5%, the forecasts that have demonstrated greater accuracy than the ARIMA model's prediction were selected for testing the forecast encompassing. The concept behind forecast combinations is that even a less accurate forecast can offer valuable information not present in the better forecast. As a result, the combined forecast can potentially achieve even greater accuracy.

CHAPTER 5

Results

Table 5.1 provides an overview of the model's performance metrics alongside the baseline performance - representing the RMSE if predicted the mean value of the federal funds rate. In addition to the RMSE metric, the p-values obtained from the Diebold-Mariano test, were also included in the table.

Based on the test set results, all models outperform the baseline, with the highest RMSE (0.577) being less than 50% of the baseline RMSE (1.175). Among the models, ARIMA exhibits the lowest forecast accuracy regarding RMSE, followed by LightGBM with RMSE values of 0.577 and 0.490, respectively. Additionally, the remaining models show similar performances, with neural networks comprising the top 3 models with the lowest RMSE: GRU (0.229), LSTM (0.247), and Bi-LSTM (0.252). Among the all models, the GRU model appears to produce the most accurate forecast, as it has the lowest RMSE.

Table 5.1.:*Train, validation and test performance of models and baseline (RMSE)*

Models	RMSE Train	RMSE Validation*	RMSE Test	P-value DM
Baseline	3.409	0.099	1.175	
ARIMA	0.532	0.047	0.577	
Random Forest	0.484	0.047	0.290	0.066**
LightGBM	0.700	0.754	0.490	0.572***
LSTM	0.513	0.040	0.247	0.040*
Bi-LSTM	0.619	0.049	0.252	0.037*
GRU	0.566	0.099	0.229	0.029*

Notes: The fifth column refers to the p-values obtained from the DM test of the significance of the forecast accuracy of each method versus that of the ARIMA model. *, ** and *** indicates the significance level, 1%, 5% or 10% (respectively) that the null hypothesis is accept.
* The RMSE from validation set represents the optimized RMSE achieved during the tuning process.

The null hypothesis of the DM test posits that there is no difference in accuracy between two competing forecasts. Therefore, with a p-value greater than the significance level of 5%, do not reject the null hypothesis, suggesting no significant difference in forecast accuracy between the two models. In simpler terms, under the null hypothesis, both models perform equally well regarding forecasting accuracy. Consequently, the results are reported in Table 5.1, and indicate that the null hypothesis is rejected at a significance level of 5% for the top 3 models, signifying that their improvements relative to the ARIMA model are statistically significant. In contrast, the superior performance of the ensemble models is not statistically significant.

Figure 5.1 displays the monthly EFFR alongside the corresponding forecasts generated by each model during the test period. It is evident that the ARIMA model (red line) and LightGBM (purple line) frequently exhibit overpredictions of the actual EFFR values, aligning with their higher RMSE scores.

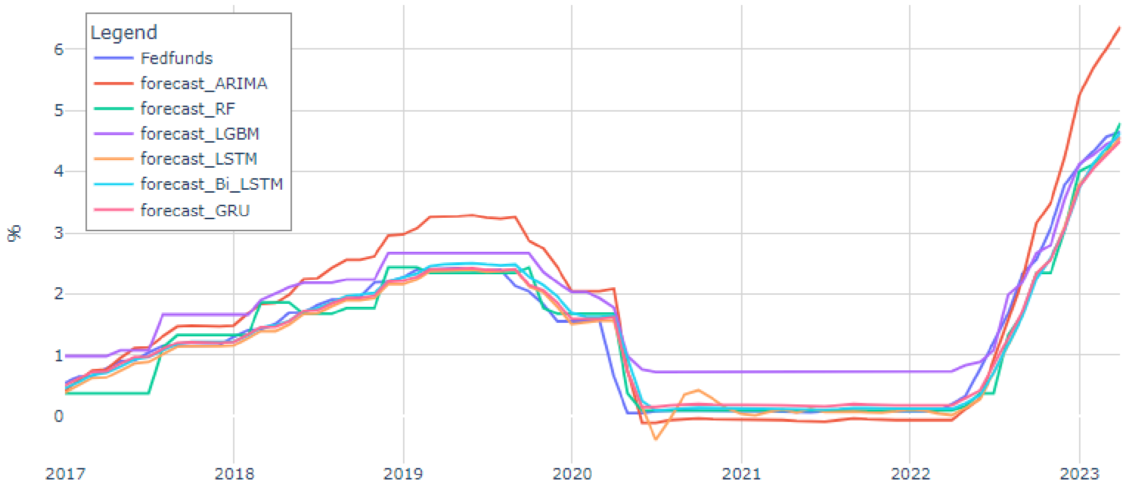

Figure 5.1.: Effective Federal Funds Rate monthly forecast (%)

According to the Figure 5.1., the last 12 months depict a steady increase in the interest rate, deviating slightly from the trend of the test set. Table 5.2 compares the top 3 best models during the test period (GRU, LSTM and Bi-LSTM) with their performance over the last 12 months to determine if they can sustain their good performance amidst the observed changes in the effective federal funds rate during this period.

As reported in table 5.2., LSTM model performance for the last 12 months decreases by 66% relative to its test period performance, while GRU and Bi-LSTM models seem to have more ability to maintain their performance over the last year (their RMSE reduces around 26% and 27%, respectively). Also, Bi-LSTM and GRU increased their RMSE percentage difference against ARIMA when isolating their forecasts only for the last 12 months, whereas LSTM tends to decrease it.

Moreover, upon closer inspection of the best model's forecast for the past 12 months (GRU) – March 2022 to Fev 2022 (depicted in Figure 5.2), it attains an RMSE of 0.289. Although its forecasting performance over this period is worse than in the test period (as shown in Table 5.2), the GRU model further reinforces its superiority over the benchmark model forecast in terms of RMSE, shifting from a reduction of 60.3% to 69.1%.

Overall, concerning both the test period and the last 12 months, the neural network models demonstrate an ability to reduce the RMSE by approximately 56%-60% and 56%-69%, respectively, compared to the ARIMA benchmark, with the GRU model emerging as the top performer in both periods.

Table 5.3: *Model's performance over the last 76 (test) and 12 months, changes in RMSE, and comparisons of RMSE with the Benchmark model*

Models	RMSE test	RMSE last 12 months	RMSE change (%)	RMSE test reduction against ARIMA (%)	RMSE last year reduction against ARIMA (%)
ARIMA	0.577	0.934	61.9	-	-
LSTM	0.247	0.412	66.8	57.2	55.9
Bi-LSTM	0.252	0.321	27.4	56.1	65.6
GRU	0.229	0.289	26.2	60.3	69.1

Notes: The RMSE values for all models presented in this table were obtained from models trained with all available information known up to the test period considered.

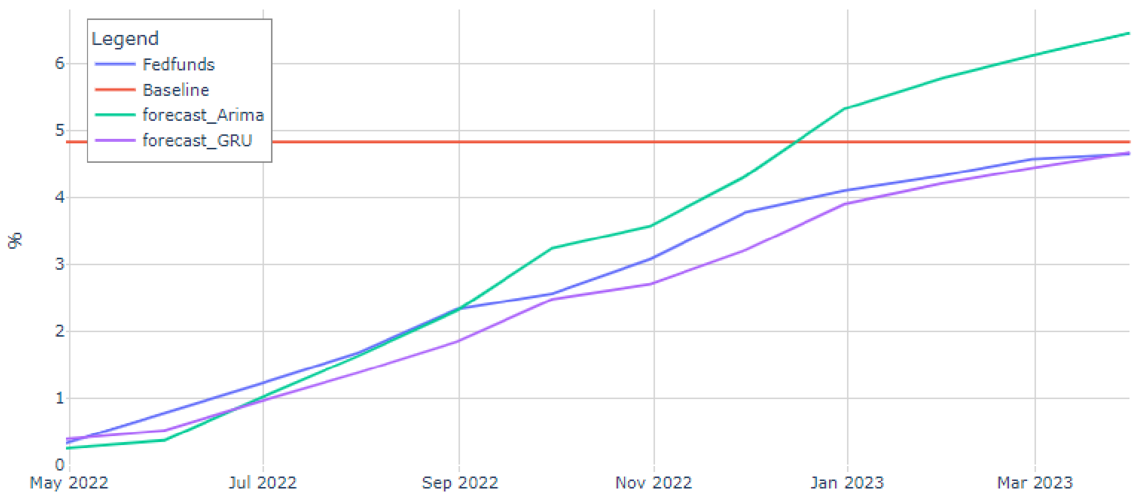

Figure 5.2.: Best model forecast, Baseline forecast and ARIMA forecast for the last 12 months

Furthermore, as explained in the previous section, we will investigate whether combining the forecasts from the models that demonstrated greater accuracy (according to the DM test) adds value to individual forecasts, as previously studied by Clark & McCracken (2008). Formally, the approach presented by Romer & Romer (2008) was followed and estimated a regression of the form:

$$y_t = a_i + \beta_i S_{it} + \alpha_i P_{it} + e_{it} \tag{13}$$

Where y_t is the actual value of the effective federal funds rate at time t, S_{it} and P_{it} are the forecasts of two of the three candidate models (LSTM, Bi-LSTM, or GRU) and e_{it} is the error term. The main interest is whether the coefficients estimates β_i and α_i are statistically significant because forecast encompassing is examined via the significance of these coefficients.

Regardless of the forecasting combination used, the residuals from equation (13) exhibit non-constant variance, a condition known as heteroscedasticity. This can potentially pose issues in interpreting the model's coefficients since the ordinary least squares (OLS) estimator assumes homoscedasticity (constant variance of residuals). Therefore, I employ weighted least squares (WLS) method since the ordinary least squares assumption of constant variance in the errors is violated.

The results are presented in Table 5.3. One of the most surprising findings pertains to the GRU forecast coefficients, both of which are negative. This implies that when attempting to predict the EFFR using either the LSTM or the Bi-LSTM forecasts and seeking to enhance this prediction by incorporating or combining forecasts from the GRU model, it is advisable to move away from the GRU forecasts rather than towards them. This may imply that the predictions of LSTM and BI LSTM models are overestimating the true value of the federal funds rate, as adding the forecast contribution of the GRU tends to reduce the predicted value of the interest rate.

Another noteworthy discovery is that the forecasts generated by LSTM and Bi-LSTM can account for only 48.6% of the variance in the actual EFFR.

Table 5.3.: *Role of LSTM, Bi-LSTM and GRU forecasts in predicting the actual values of the EFFR*

Combination	LSTM coefficient	Bi-LSTM coefficient	GRU coefficient	R^2 (%)
LSTM & Bi-LSTM	1.32	-0.28*	-	48.6
LSTM & GRU	6.25	-	-5.37	93.8
Bi-LSTM & GRU	-	3.13	-2.26	85.5

Notes: These results are obtained with WLS estimation using the training data. Also, the statistically significance of coefficients is identified by the asterisks, as follows: *, **, *** indicates that there is evidence that the corresponding forecast is not statistically significant to explain the variation in the actual values, assuming a significance level of 1%, 5% and 10% respectively. Whenever there is no asterisk, it means rejection of the null hypothesis for any significant level, and therefore the forecast is statistically significant.

However, all the estimates of the coefficients are statistically significant at a 5% level, which supports the hypothesis that forecast combination can enhance performance. This suggests that there is no redundancy in adding information from these forecasts.

Table 5.3 reports the RMSE from all possible forecast combinations under consideration. Two different strategies were considered: one using an equal weighting strategy and the other using a weighted average by attributing a higher weight to the more accurate forecast for it to have a greater contribution to the final combined forecast.

Consider two alternative sets of forecasts for a variable, denoted as f_1 and f_2. Suppose that f_1 outperformed f_2 in the validation set, with an RMSE of 0.2 compared to an RMSE of 0.4. One approach combines these two forecasts with equal weights, 50% each. An alternative approach is to give more weight to f_1, as it has a lower RMSE. The weights, denoted as α_1 and α_2, are defined as follows:

$$\alpha_1 = \frac{RMSE_2}{(RMSE_2 + RMSE_1)} \tag{14}$$

$$\alpha_2 = 1 - \alpha_1 \tag{15}$$

$$F = \alpha_1 * f_1 + \alpha_2 * f_2 \tag{16}$$

where F is the combined forecast.

We merely utilized the RMSE weights obtained from each individual forecast ($RMSE_2$ and $RMSE_1$) in the validation set to assign them as the new weights for the opposing forecasts in the combined forecast. We are ensuring that the contribution of an individual forecast with a higher RMSE will not only not have an equivalent impact on the combined forecast compared to an individual forecast with a lower RMSE but also will have a lesser impact.

However, there is dependency regarding this approach. The choice of weights depends on the proportion of each RMSE over the validation data for each individual forecast. It can be summarized that if this proportion does not follow the same distribution over the test data, the selection of these weights may not yield any improvement. This is why no enhancement is observed when combining the GRU forecasts with either of the other two forecasts (compare the individual RMSE values from Table 5.1 with the combined RMSE reported in Table 5.4).

According to the results presented in Table 5.1, the GRU model exhibited the highest RMSE (0.099) in the validation set compared to the RMSE of the LSTM and Bi-LSTM models (0.04 and 0.049, respectively). However, in the test set, the situation reversed, with the GRU's forecast proving to be the most accurate.

Table 5.5.*: Forecast Encompassing RMSE over the test data*

Combination	RMSE equal weight	RMSE weighted average
LSTM & Bi-LSTM	0.244	0.243
LSTM & GRU	0.232	0.237
Bi-LSTM & GRU	0.238	0.244

Table 5.5. reports the forecast encompassing RMSE overt the test data for the three possible combinations. Overall, forecast combinations allow LSTM and Bi-LSTM to reduce RMSE by approximately 1% to 6% and 3% to 6%, respectively, which is not very significant. And yet, the combination of forecasts failed to outperform the individual forecast from the GRU. This poor performance of the forecast combination had already been documented in previous studies by Hinterlang, 2019 and Sarno et al., 2005.

CHAPTER 6

Conclusion

This thesis aimed to select the most important variables to predict the EFFR using several feature selection techniques and the Granger Causality test. From the set of variables that appear to better explain EFFR behavior, financial and labor market indicators are the two main contributing features. Moreover, according to the Granger Causality test, past and current information about the Price-Earnings Ratio from S&P's Composite Common Stock is not statistically significant to predict the EFFR, assuming a 1% significant level.

Furthermore, the performance of some popular ML algorithms was evaluated in obtaining accurate forecasts of the Effective Federal Funds Rate. For that, several models were estimated over the 1960-2016 period with the corresponding optimal hyperparameters and corresponding optimal variables set. By comparing the forecasts obtained from these models over the 2017- Fev 2023 period with the accuracy of a simple average forecast for the same period, we found that all the models could produce more accurate forecasts (according to the RMSE).

The results also suggest that the deep learning algorithms, namely LSTM, Bi-LSTM, and GRU, outperformed Random Forest (RF) and LightGBM. In particular, GRU is considered the best model, since it has the lowest RMSE. Furthermore, when comparing forecast accuracy using the Diebold-Mariano (DM) test, the deep learning algorithms significantly outperformed the naive ARIMA benchmark regarding forecasting performance. In contrast, the highest accuracy of Random Forest and LightGBM forecasts against the benchmark was not statistically significant.

It was also analysed if the top 3 models were able to maintain or even increase their predictive abilities over the last 12 months, since it is a period that deviates quite a bit from the overall behavior of the test period. It was found that, all models reduced their performance in terms of RMSE, particularly LSTM experienced the largest decrease of it. However, GRU and Bi-LSTM were able to improve their performance against the benchmark model, while once again LSTM did not.

Moreover, the combination of forecasts to reduce RMSE does not appear to make a substantial contribution in this regard, as the individual forecasting performance of the GRU model remains higher than its combined performance with the others.

Nevertheless, it is crucial to recognize this study's limitations due to the absence of prior research on forecasting the effective federal funds rate. This limits the opportunity to establish a benchmark for comparing the results in this paper, such as evaluating whether the RMSE falls within the typical range observed in the literature. Another limitation is the restricted number of observations, as some ML algorithms can be sensitive to the size of the dataset. Certain algorithms may require a minimum amount of data to function effectively or to learn meaningful patterns.

Despite the acknowledgment of the limitations of this study, the obtained results strongly recommend utilizing machine learning (ML) algorithms, especially deep learning (DL) algorithms, as complementary tools to assist policymakers in making more sustainable and effective monetary policy decisions. These algorithms are also valuable for investors and financial institutions, as they can help inform decision-making processes and capital allocation, for example. Furthermore, consumers should not overlook their relevance, as the effective federal funds rate indirectly impacts their finances and choices related to borrowing, saving, and investing. Additionally, this dissertation contributes to the growth of studies related to the application of ML algorithms in forecasting the Effective Federal Funds Rate, as it is a topic that has received very little attention in the literature.

References

Board of Governors of the Federal Reserve System (US). (2016). The Federal Reserve System Purposes and Functions: Conducting Monetary Policy. Washington, DC: Board of Governors.

Board of Governors of the Federal Reserve System. (2018). The Federal Reserve System Purposes and Functions: Monetary Policy Principles and Practice. https://www.federalreserve.gov/monetarypolicy/policy-rules-and-how-policymakers-use-them.htm

Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324

Buckley, N. (2022, December 30). FT writers' predictions for the world in 2023. The Financial Times. Retrieved from https://www.ft.com/content/9784cc74-1193-4e1b-bf61-8ecaf19f569e

Chan, Y. H. (2003). Biostatistics 104: Correlation Analysis. Signapure Medical Journal, Vol 44(12) : 614-619.

Cho, K., Van Merrienboer, B., Bahdanau, D. and Bengio, Y.. On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259, 2014.

Choi, R., Coyner, A., Kalpathy-Cramer, J., Chiang, M., & Campbell, J. (2020). Introduction to Machine Learning, Neural Networks, and Deep Learning. Trans Vis Sci Tech, 9(2).

Chu, B. & Qureshi, S. (2022). Comparing Out-of-Sample Performance of Machine Learning Methods to Forecast US GDP Growth. Computational Economics.

Clark, Todd E. and McCracken, Michael W. (2008). Improving Forecast Accuracy by Combining Recursive and Rolling Forecasts. FRB of Kansas City Working Paper No. RWP 04-10, FRB of St. Louis Working Paper No. 2008-028A, Available at SSRN: https://ssrn.com/abstract=615122

Clarida, R., Gertler, M., & Galí, J. (1998). Monetary policy rules in practice: Some international evidence. European Economic Review, 42(6), 1033-1067.

De Grauwe, P., & Ji, Y. (2021). On the Use of Current or Forward-Looking Data in Monetary Policy: A Behavioural Macroeconomic Approach. CESifo Working Paper Series No 8853.

Diebold, F., Mariano, R. (1995). Comparing Predictive Accuracy. Journal of Business & Economic Statistics, vol. 13, issue 3, 253-63.

El-Shagi, M., & Jung, A. (2015). Have minutes helped markets to predict the MPC's monetary policy decisions? European Journal of Political Economy, 39, 222-234.

Fair, R. C. (2001). Actual Federal Reserve Policy Behavior and Interest Rate Rules. Economic Policy Review, 7(1), 12.

Hayo, B., & Neuenkirch, M. (2010). Do Federal Reserve communications help predict federal funds target rate decisions? Journal of Macroeconomics, 32(4), 1014-1024.

Hinterlang, N. (2019). Predicting Monetary Policy Using Artificial Neural Networks. Deutsche Bundesbank Discussion Paper No. 44/2020.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Holzinger, A. (2019). Introduction to Machine Learning & Knowledge Extraction (MAKE). Machine Learning and Knowledge Extraction, 1(1), 1-20.

Ho, T. K. (1995). Random decision forests. Proceedings of 3rd International Conference on Document Analysis and Recognition (pp. 278-282 vol.1). Montreal, QC, Canada: IEEE.

Hu, L., & C. B. Phillips, P. (2004). Dynamics of the Federal Funds Target Rate: A Nonstationary Discrete Choice Approach. Journal of Applied Econometrics, 19, 851-867.

Hubert, P. (2015). The Influence and Policy Signalling Role of FOMC Forecasts. Oxford Bulletin of Economics and Statistics, 77(5), 655-680.

International Monetary Fund (IMF). (2022, March 3). Monetary Policy and Central Banking. https://www.imf.org/en/About/Factsheets/Sheets/2023/monetary-policy-and-central-banking

International Monetary Fund (IMF). (2022, October 11). World Economic Outlook: Countering the Cost-of-Living Crisis. https://www.imf.org/en/Publications/WEO/Issues/2022/10/11/world-economic-outlook-october-2022#:~:text=COUNTERING%20THE%20COST%2DOF%2DLIVING%20CRISIS,-OCTOBER%202022&text=This%20is%20the%20weakest%20growth,to%204.1%20percent%20by%202024

Kauppi, H. (2012). Predicting the Direction of thre Fed's Target Rate. Journal of Forecasting, 31, 47-67.

Kim, H., Jackson, J., & Saba, R. (2009). Forecasting the FOMC's Interest Rate Setting Behavior: A Further Analysis. Journal of Forecasting, 145-165.

Lapp, J. S., & Pearce, D. K. (2003). The Predictability of FOMC Decisions: Evidence from the Volcker and Greenspan Chairmanships. Southern Economic Journal, 70(2), 312-327.

Lee, K., & Morley, J. (2015). The Meta Taylor Rule. Journal of Money Credit and Banking, 47(1), 73-98.

Leith, C., & Wren-Lewis, S. (2009). Taylor rules in the open economy. European Economic Review, 53(8), 971-995.

Lucio, S., Thornton, D., & Giorgio, V. (2005). Federal Funds Rate Prediction. Journal of Money, Credit, and Banking, 37(3), 450-471.

Maleki, F., Ovens, K., Najafian, K., Forghani, B., Reinhold, C., & Forghani, R. (2020). Overview of Machine Learning Part 1 Fundamentals and Classic Approaches. Neuroimag Clin N Am, 17-32.

Masini, R., Medeiros, M., & Mendes, E. (2021). Machine Learning Advances for Time Series Forecasting. Journal of Economic Surveys, 1-36.

Mitchell, T. M. (1997). Machine Learning. McGraw-Hill Science/Engineering/Math.

Molodtsova, T., Nikolsko-rzhevskyy, A., & Papell, D. (2011). Taylor Rules and the Euro. Journal of Money, Credit and Banking, 535-552.

Monokroussos, G. (2011). Dynamic Limited Dependent Variable Modelingand U.S. Monetary Policy. Journal of Money, Credit and Banking, 43(2/3), 519-534.

Pakko, M. R. (2005). On the information content of asymmetric fomc policy statements: evidence from a taylor-rule perspective. Economic Inquiry, 467-697.

Pauwels, L. L., & Vasnev, A. L. (2017). Forecast combination for discrete choice models: predicting FOMC monetary policy decisions. Empirical Economics **52**, 229-254.

Piazzesi, M. (2005). Bond Yields and the Federal Reserve. Journal of Political Economy, 113(2), 311-344.

Qifang, B., Goodman, K., Kaminsky, J., & Lessler, J. (2020). What is Machine Learning? A Primer for the Epidemiologist. American Journal of Epidemiology, 188(12), 2222-2239.

Quah, J., & Hemamalini, V. (2008). Tracking Federal Funds Target Rate Movements Using Artificial Neural Networks. 2008 IEEE Conference on Soft Computing in Industrial Applications, Muroran, Japan, 246-251.

Radovic, M., Ghalwash, M, Filipovic, N. & Obradovic, Z. (2017). Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. BMC Bioinformatics.

Romer, C. D., & Romer, D. H. (2008). The FOMC versus the Staff: Where Can Monetary Policymakers Add Value? AMERICAN ECONOMIC REVIEW, VOL. 98, NO. 2, pp. 230-35.

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of Research and Development, 3(3), 210-229.

Sarno, L., Thornton, D. L., & Valente, G. (2005). Federal Funds Rate Prediction. Journal of Money Credit and Banking, 449-471.

Seibert, A., Sirchenko, A., & Müller, G. (2021). A model for policy interest rates. Journal of Economic Dynamics and Control, 124.

Svensson, L. E. (2002). What is Wrong with Taylor Rules? Using Judgment in Monetary Policy through Targeting Rules. Working Paper No. 9421. National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w9421/w9421.pdf

Svensson, L. E. (2020). What Rule for the Federal Reserve? Forecast Targeting. International Journal of Central Banking, 16(6), 39-95.

Taylor, J. B. (1993). Discretion versus policy rules in practice. Carnegie-Rochester Conference Series on Public Policy, 195-214.

US Bureau of Labor Statistics. (2023, September 17). All Employees, Goods-Producing/All Employees, Total Nonfarm . Retrieved from FRED, Federal Reserve Bank of St. Louis: https://fred.stlouisfed.org/series/PAYEMS

Van den Hauwe, S., van Dijk, D., & Paap, R. (2013). Bayesian forecasting of federal funds target rate decisions. Journal of Macroeconomics, 37, 19-40.

Yan-jie, J., Liang-peng, G., Xiao-shi, C., & Wei-hong, G. (2017). Strategies for multi-step-ahead available parking spaces forecasting based on wavelet transform. Journal of Central South University, 24, 1503−1512

Yu, L., Liu, H. (2003). Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution. Machine Learning, Proceedings of the Twentieth International Conference. Washington: DBLP.

Zeng, Z. Z. (2015). A Mixed Feature Selection Method Considering Interaction. Mathematical Problems in Engineering.

Appendix A

Table A1

New variable's description

Variables	Description
10-Year Real Interest Rate	Nominal interest rate adjusted for inflation over 10 years
Short-term Inflation Expectation [MICH]	Measure of short-term (12 months) inflation expectations based on consumer surveys and economic experts' opinions.
10-Year Expected Inflation	Anticipated average annual inflation rate over a ten-year period
Inflation	Rate at which the general price level of goods and services in an economy rises over a period of time
Brave-Butters-Kelley Indexes [*BBKMGDP]*	Brave-Butters-Kelley Real Gross Domestic Product growth is indexed to the quarterly estimates of real GDP growth from the US Bureau of Economic Analysis and consists of three components: cycle, trend, and irregular components[11]

[11] The authors created a specialized statistical model known as a "mixed-frequency collapsed dynamic factor model." This model was designed to analyze and estimate the unobservable monthly changes in the quarterly growth rate of the U.S. real GDP by utilizing a wide range of 500 monthly time series data points. For detailed information https://www.chicagofed.org/publications/chicago-fed-letter/2019/422

Table A2

Top 5 variables chosen by each feature selection method

Variables	RFE	IWFS	FCBF	MRMR	SFSF	SFSB	Tree	VIF	Total
CP3Mx	x	x	x	x	x	x	x		7
TB3MS		x	x	x			x		4
TB6MS		x	x	x			x		4
COMPAPFFx	x				x	x			3
S&P PE ratio							x	x	2
MICH						x		x	2
BBKMGDP						x		x	2
GS1		x							1
BAAFFM		x							1
CUSR0000SAS			x						1
CPIMEDSL			x						1
IPCONGD				x					1
NONBORRES				x					1
RPI					x				1
USGOOD	x								1
PAYEMS	x								1
SRVPRD	x								1
DPCERA3M086SBEA					x				1
W875RX1					x				1
REAINTRATREARAT10Y						x			1
CPALTT01USM657N								x	1
VIXCLSx								x	1

Notes: The RF selected only 4 features. In the other methods, I had to predefine an initial number of features to be selected. The last column corresponds to the number of times each feature was chosen to be part of the top 5 features.

Table A3

Final variables set description

Variables	Description
CP3Mx	3-Month AA Financial Commercial Paper Rate
TB3Ms	3-Month Treasury Bill
TB6Ms	6-Month Treasury Bill
COMPAPFFx	3-Month Commercial Paper Minus FEDFUNDS
S&P PE ratio	S&P's Composite Common Stock: Price-Earnings Ratio
BBKMGDP	Brave-Butters-Kelley Real Gross Domestic Product growth is indexed to the quarterly estimates of real GDP growth from the US Bureau of Economic Analysis and consists of three components: cycle, trend, and irregular components
Short-term Inflation Expectation [MICH]	Measure of short-term (12 months) inflation expectations based on consumer surveys and economic experts' opinions.
USGOOD	Goods-Producing Industries that measures the number of US workers in industries involving the production of tangible goods or products. They typically include activities related to manufacturing, construction, mining and agriculture (measured as thousands of persons)
SRVPRD	Service-Providing Industries that measures the number of US workers in industries that provide intangible services rather than tangible goods. The service sector is diverse and includes a wide range of activities, such as healthcare, finance, education, retail, hospitality, and professional services (measured as thousands of persons)
PAYEMS	Total nonfarm that measures the number of US workers in the economy excluding proprietors, private household employees, unpaid volunteers, farm employees and the unincorporated self-employed (measured as thousands of persons)

Table A4

P-values from Granger Causality test

Y \ X	Fedfunds	Srvprd	Cp3mx	Usgood	Bbkmgdp	S&P pe ratio	Mich	Compaff	Tb3mx	Tb6ms	Payms
Fedfunds	**1**	0	0	0	0	0.013	0.002	0	0	0	0
Srvprd	0.001	**1**	0	0	0	0.002	0	0.013	0.150	0.012	0
Cp3mx	0	0	**1**	0	0	0.423	0.054	0	0	0	0
Usgood	0.0001	0	0	**1**	0	0.090	0	0	0.005	0	0
Bbkmgdp	0.468	0	0	0	**1**	0.292	0.011	0	0.639	0.301	0
S&P pe ratio	0	0	0.007	0	0.003	**1**	0	0	0	0	0
Mich	0.083	0	0.015	0.142	0.021	0.024	**1**	0.020	0.270	0.241	0
Compapff	0.638	0	0.047	0.271	0.014	0.152	0.072	**1**	0.343	0.789	0
Tb3mx	0	0.010	0.020	0.002	0.002	0.101	0.030	0.080	**1**	0	0.004
Tb6mx	0	0	0	0	0	0.554	0.038	0.026	0	**1**	0
Payms	0.001	0	0	0	0	0	0	0.005	0.114	0.009	**1**

Notes: These results are obtained considering a maximum 15 lags (five quarters)

Table A5.

Correlation Coefficients

Y \ X	Fedfunds	Srvprd	Cp3mx	Usgood	Bbkmgdp	S&P pe ratio	Mich	Compaff	Tb3mx	Tb6ms	Payms
Fedfunds	**1**										
Srvprd	-0.841*	**1**									
Cp3mx	0.996*	-0.838*	**1**								
Usgood	0.744*	-0.563*	0.753*	**1**							
Bbkmgdp	0.155*	-0.229*	0.143*	0.174*	**1**						
S&P pe ratio	-0.473*	0.424*	-0.474*	-0.187*	0.057	**1**					
Mich	0.410*	-0.295*	0.421*	0.068	-0.196*	-0.450*	**1**				
Compapff	-0.404*	0.350*	-0.330*	-0.175*	-0.197*	0.169*	-0.023	**1**			
Tb3mx	0.996*	0.010	0.995*	0.756*	0.175*	-0.465*	0.39*	-0.377*	**1**		
Tb6mx	0.995*	-0.846*	0.996*	0.749*	0.171*	-0.468*	0.40*	-0.351*	0.999*	**1**	
Payms	-0.802*	-0.850*	-0.797*	-0.472*	-0.222*	0.428*	-0.31*	0.351*	-0.806*	-0.811*	**1**

Notes: * indicates correlation significant at a 1% level. None asterisk means not significant at any significance level.

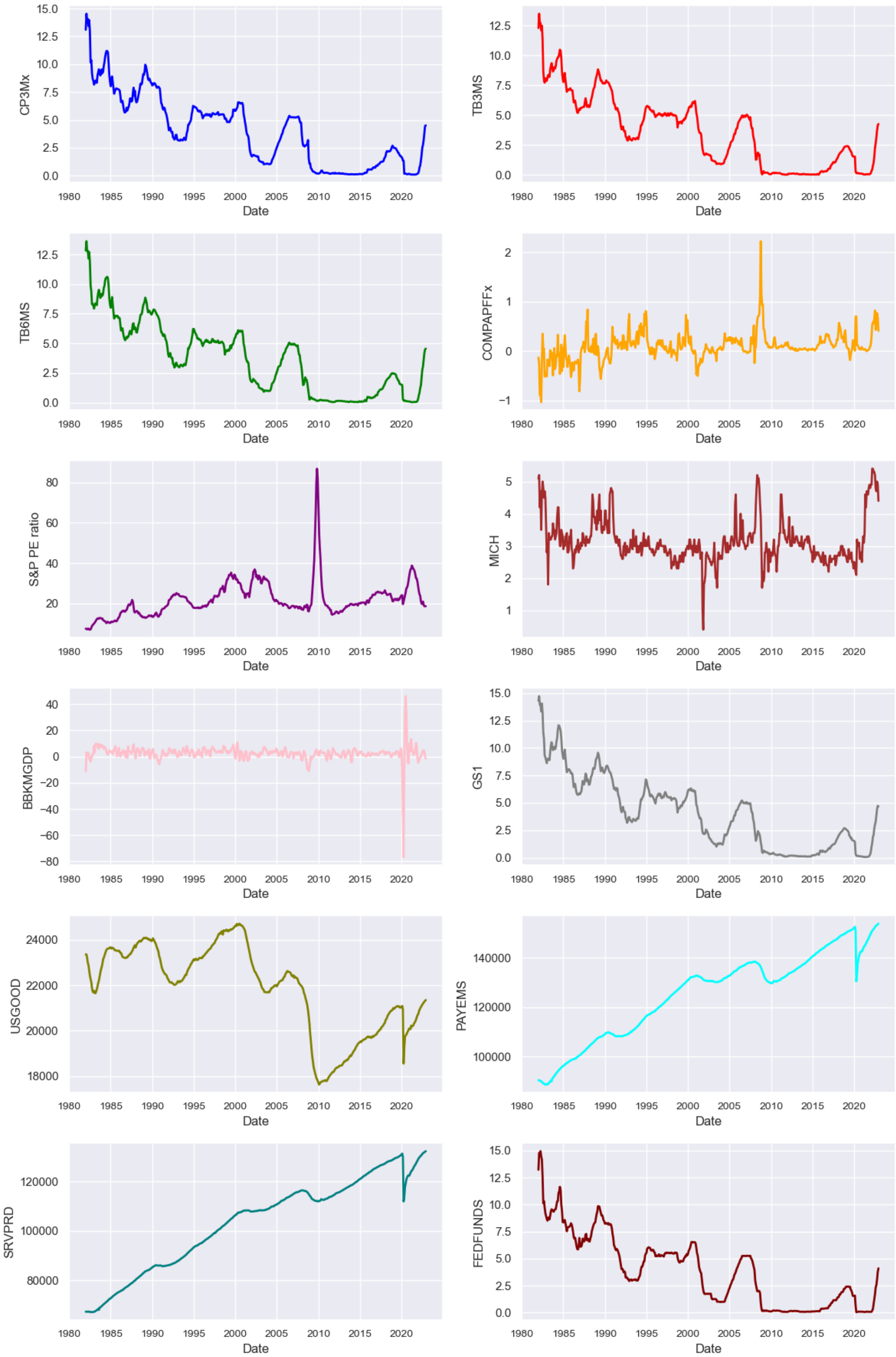

Figure A2: Time series data for January 1978 to February 2023 period

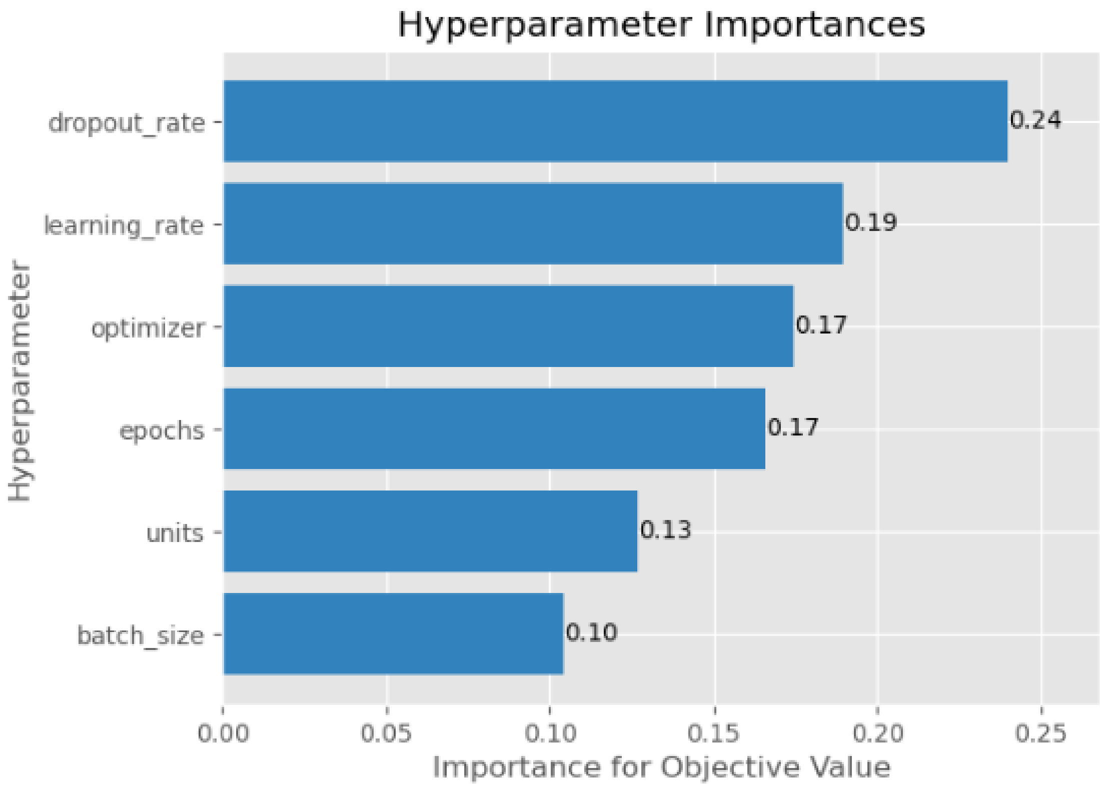

Figure A3: Hyperparameter importance analysis for GRU model

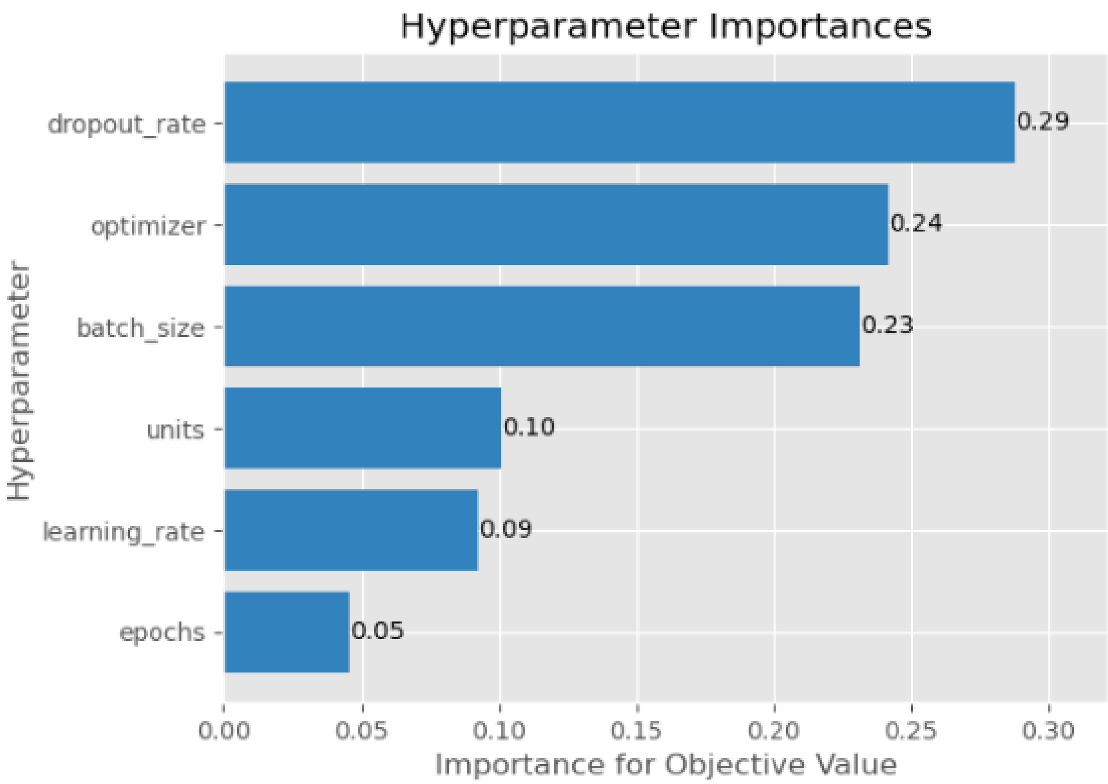

Figure A4: Hyperparameter importance analysis for LSTM model

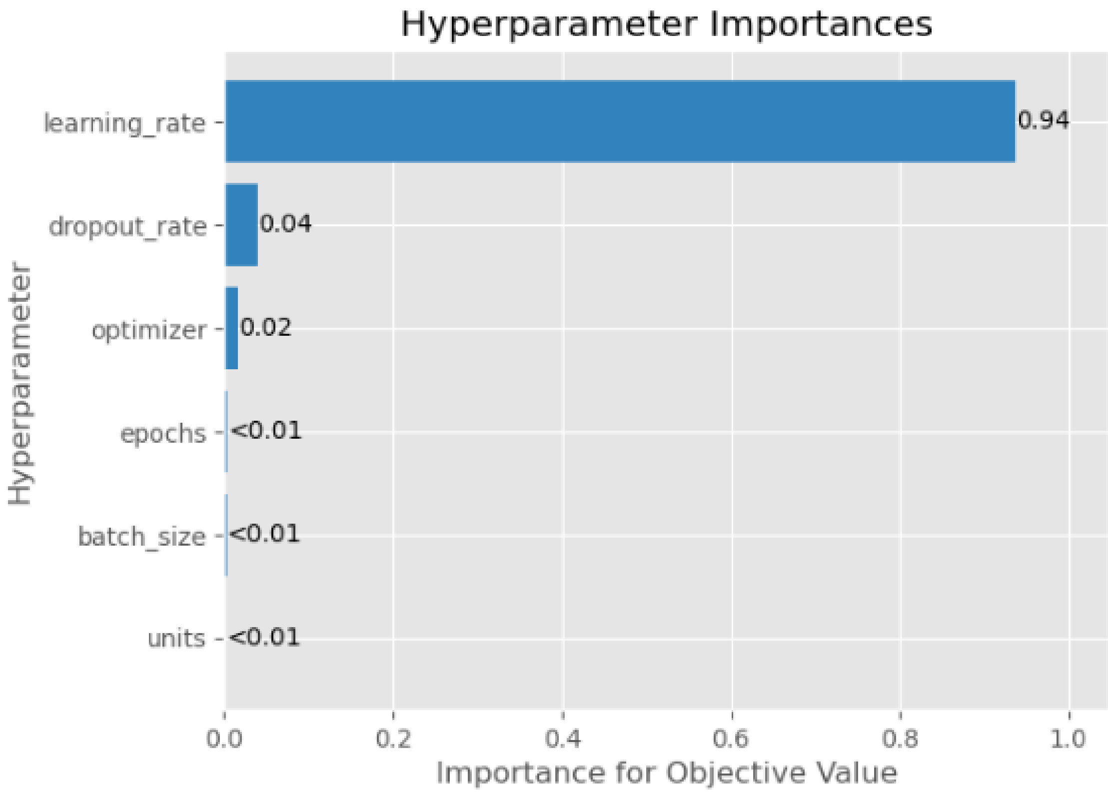

Figure A5: Hyperparameter importance analysis for Bi-LSTM model