Repo Rate Modelling based on Financial and Economic Factors

MSc Research Project
Data Analytics

Himanshu Rathee
Student ID: 20132689

School of Computing
National College of Ireland

Supervisor: Dr. Rashmi Gupta

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Himanshu Rathee
Student ID:	20132689
Programme:	Data Analytics
Year:	2021
Module:	MSc Research Project
Supervisor:	Dr. Rashmi Gupta
Submission Due Date:	16/8/2021
Project Title:	Repo Rate Modelling based on Financial and Economic Factors
Word Count:	000
Page Count:	25

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	16th August 2021

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Repo Rate Modelling based on Financial and Economic Factors

Himanshu Rathee
20132689

Abstract

Interest rate is the most important financial variable that helps in determining the macro and microeconomic policy of a country. Therefore predicting any change in the value of interest rates become crucial for the government and various other financial institutions. This research aims to forecast the interest rates for India and analyze various factors affecting the change in the interest rates for better forecasting. A novel approach of using sentiment analysis of tweets of various users about the economic situation along with various financial and economic variables that help in predicting interest rates has been used. Four machine learning models namely: Vector autoregressive model, Long short term memory, Sequential neural network, and Multilayer perceptron neural network have been compared based on evaluation metrics such as root mean square error, mean absolute error, and mean absolute error percentage. The analysis shows that the machine learning model with input as sentiment analysis along with various financial variables outperforms the models with input as various financial variables. The best model is the Multilayer perceptron neural network model trained with the novel approach and has a mean absolute error percentage of 1.76%.

1 Introduction

1.1 Background and Motivation

The interest rate is among the most essential tools for attaining the fiscal policy goals of financial institutions. This provides the foundation for many other financial sector operations, such as managing risk, investment management, speculating, product creation, etc. Consequently, preserving demand balance and excessive hikes in interest rates (Jie et al.; 2008) is the important factor with in financial sector. Financial analysts, academics, and regulators were all interested in the amount of risk associated in the interest rate, that should plateau when there are no temporary disruptions. However, several structural developments have raised concerns regarding the implications of these developments on world economies in sustaining actual rate balance (Yellen et al.; 2015). In general, a central bank as well as other regulators set a fiscal policy price that acts as a benchmark for all localized interest rates and incorporates maturities for providing risk-free mortgages in various denominations (Harrison et al.; 2005). Interest rates have been at the vanguard of the public conversation for many years, as they have a very powerful effect on economic development. The levels at which interest rates are kept are controlled by many judgments, such as how much individuals are to spend and save, companies that decide how much they are to invest, and government evaluating different financial and fiscal policies. The measure of economic interest rates has a wide range of consequences. This is why the interconnections, linkages and correlations between interest rate and numerous other economic data (Pejović and Karadžić; 2020a) are examined by different modelling and prediction approaches.

Artificial intelligence and other optimization techniques are utilized to handle several of the major funding issues such as interest rate fluctuations predicting. Machine learning is all the more essential because of the complicated and nonlinear character of the interest rates. Interest rates are a significant component of government policy on macroeconomic policy and hyperinflation (Ju et al.; 1997; Enke and Thawornwong; 2005).

In recent decades, various financial theories based on complex mathematical and market equilibrium theory have been suggested for precise estimate and monitoring of interest rate fluctuations. Furthermore, several stochastic process theories are often used in the literature to capture interest rate fluctuations. The majority of these models are economic theories, but they do not have a role in determining rate fluctuations in the disaster period such as a pandemic (Vasicek; 1977a). To correctly analyze and predict interest rate fluctuations, a more adaptable and accurate model is required. Therefore, various financial and economic variables that determine the changes in the interest rates have to be visualized. However, sometimes the pressure from people as well as government instigates the changes in the interest rates which makes it interesting to analyze the role of people's sentiment regarding the policies and if it helps in stimulating the forecasting of interest rates. Therefore the focus of this research is to study four state-of-the-art models and determine various factors that contribute to the forecasting of interest rates.

1.2 Research Question

Through the following research questions, the goal of this study is to gain a better knowledge of the study of predicting interest rates.

Can machine learning techniques be used to forecast a country's repo rates based on different financial and economic factors?

Sub-questions

1. *How much effect does each element have on the fluctuations in exchange rates?*
2. *Is it true that using diverse inputs enhance the efficiency of rate forecasting?*

1.3 Objective

The primary goal of this research is to predict the central banks' interest rates and analyze the impact of various financial and economic factors on the projection. Furthermore, the shift in interest rates is significant during a crisis. Incorporating emotional stakeholder analysis with deep learning may also result in more accurate interest rate projections.

1.4 Contribution

This study aids in the forecasting of one of the most significant financial and economic factors, the interest rate at which central banks lend money to retail and government banks. Interest rates are important because they influence nearly every micro and macroeconomic choice made by a government, as well as every investment opportunity. People frequently make decisions by anticipating the future, thus understanding the current period's elements that may impact the future is critical.

As a result, this research contributes by analyzing different economic and financial elements that impact interest rate predictions, as well as the function of people's opinion about a country's economic situation in influencing interest rate fluctuations.

1.5 Report Structure

This research report is broken into seven sections, each of which focuses on a distinct analysis of research. Section 1 presents the research topic and demonstrates that the forecasting of interest rates are important. Section 2 covers prior study conducted in several fields on this topic. The approach to finish the research job is focused on in section 3. Section 4 outlines the research design specifications. The research project is conducted in Section 5 and the whole procedure of the research problems is completed. Section 6 includes the methods for evaluation and the outcomes of several models of machine learning. Section 7 concludes the study effort and proposes scope for the future.

2 Related Work

Literature analysis is a critical component of any research since it helps to a better knowledge and awareness of the many techniques and tactics employed in the evaluation of the subject under consideration. There seems to be a lot of study in the financial industry, both currently and in past eras, and rates of interest are an important financial metric. Because interest rates are so important in economics and finance, both intellectual and empirical approaches are employed to anticipate interest rates.

A detailed grasp of interest rates, the influence of other fundamental ratios on interest rates, and the many models used in predicting becomes critical to a deeper understanding of interest rate prediction. As a result, this literature evaluation has been divided into the following three sections to provide a concise and influential information discovery regarding interest rate forecasting.

1. Theory Based Models
2. Mathematical Models
3. Machine Learning Models
   (a) Basic Models
   (b) Deep Learning Models

2.1 Theory Based Models

Different theoretical analysis approaches were attempted to assess advances in interest rate curves, but four approaches have been well acknowledged.

Unbiased expectations theory (Hicks et al.; 1975) depicts diverse investors' predictions regarding the future inflation and interest rates. The liquidity preference theory (Hicks et al.; 1975), which enhances the expectation theory by taking short-term rate liquidity into consideration. The market segmentation hypothesis (Gordon et al.; 2003), which takes into consideration a rule that restricts purchasers and borrowers from entering a new market when they have already left, even if existing market circumstances are adverse. The preferred habitats hypothesis (Gordon et al.; 2003),addresses the favorite segments of the market for lenders and investors, and moreover their capacity to migrate across submarkets if there are significant variations in returns across segments.

These theoretical approaches assist to a greater grasp of interest rates and present an indication of the variables that influence rate fluctuations.

2.2 Mathematical Models

Four important literature parts utilized for forecasting interest rates from factual and mathematical perspective are given below.

The gradient of the yield curve is a crucial economic variable in the first portion of models for forecasting GDP growth and inflation expectations (Estrella and Mishkin; 1997). Several non-linear and linear empirical models in this area believe the short-term bond rate to be a significant characteristic in forecasting inflationary and other economic factors, such as interest rates, manufacturing output, recession indicators, and so on (Bernanke et al.; 2005). The second category of modeling is known as data-driven models, and it includes methods for matching arithmetic to spot prices. However, because their capacity to anticipate future ambiguity is limited as well as the current shape of the term structure is much more unpredictable, they are ineffective at forecasting interest rates (McCulloch; 1971). The dynamic models are indeed the third model category. Arbitration-free models and balancing models, such as the Cox Ingersoll Ross (CIR) models, are examples of this. The fourth component of models is data-driven and relies on information finding techniques (KDD). It is critical to bypass these segments' constraints and evaluate complicated nonlinear relationships between variables. In addition to various advantages, this technique addresses seasonality and the presence of fundamental gaps inside the data. As a result, the majority of previous work in this model part is devoted to forecasting interest rates (Vela; 2013).

Understanding of these parametric studies is important since it aids in the development of strategies used in current interest rate forecasting. It also clarifies the analytical approach and provides a clear grasp of which models outperform others, indicating that some of the most beneficial models would be those developed using the information finding techniques.

2.2.1 Financial Models

While Knowledge discovery is the finest and correct selection for all quantitative modeling approaches, it's also very essential to thoroughly study some of the valuation forecasting in order to acquire a comprehensive grasp of the advantages.

The theory of classical economics interest rate determination, the theory of the parity rate, the theory of liquidity choice and the classical theory of economic loanable money are the main theoretical description of exchange rates. One common feature in all of these models is that a constant element is described by the exchange rate choice. (Merton; 1973) looked at interest rates uniquely and took into account fluctuations throughout the rate in 2 components: the drift portion and the part that explains the stochastic influence known as the dissemination component. After these many groups were constructed for the drift and dispersion of lending rates, many others like (Cox et al.; 2005)(Chan et al.; 1992)(Vasicek; 1977b) presented new approaches and developed the concept of a particular element. The multiple factor concept (Gallant and Tauchen; 2001)that has a volatile variable or the mean return parameter changes with time is another progress in the approach of the solitary factor analysis. An analytical solution utilizing a multiple component model, however, isn't straightforward to obtain because the expression is complicated. While study on these models might infer that the findings may indicate some variance depending on different econometric methodologies, various markets and various model hypotheses.

Other most used models as a part of mathematical finance are the CIR model, the Chan Karolieyi Longstaff and Sandres (CKLS) model, and the Vasicek as they represent the advancements in interest rates. The model properties of the traditional Vasicek, CIR and CKLS methods are compared using (Ma Jie et al.; 2008). Empirical studies have shown that the bank rate enhancements in the CIR and Vasicek models are more suitable than the CKLS model.

In (Long-Zhen Fan and Lan-Jun Lao; 2004), an empirical comparison was carried out between various models in short interest and repo in the interbank China market. The repo rate was forecasted for 7 days with the aid of the SMP predictor. It featured several versions: Vasiceks, CIR, CIR0 as a generalized CIR model, Brennan-Schwartz, CKLS, and CKLS0 as a generalized CKLS model. The CIR0 and CKLS0 models have been found to fit the interest rate data correctly. The study also revealed that short-term volatility fluctuates with time and is composed of a steady element and a variable component. The major assumption made by such prominent interest rate theories would be that the short rate's instantaneous conditional means was moderate and steady.

These statistical methods provide a better knowledge of the characteristics and modeling approaches connected with forecasting interest rates, making them valuable for this study. This area of the literature highlights the possibility that mathematical models might be beneficial in properly predicting interest rates.

2.3 Machine Learning Models

A study of strategies utilized in the machine learning field to anticipate interest rates is critical in determining the appropriate strategy to increase performance or test the proposed goals. As a result, the following approaches have been examined.

2.3.1 Basic Machine Learning Models

With recurrent artificial neural networks (RNN), previous information can be utilised by backward looping either final output or an intermediate layer. As a consequence, the prediction of a recurrent network is impacted not only by the present input but also from the upcoming input. The prediction at period t is determined by the complete history before t, which addresses the difficulty of calculating the series-related delays (Zurada; 1992). The RNN model was employed as an alternative to standard recursive methods in a study of forward interest rate prediction. RNN accuracy is measured by calculating forward rates using a Monte Carlo simulation, and RNN has been proven to be adequate (Bouqata et al.; 1999).

Artificial neural networks are employed in (Oh and Han; 2000) to forecast interest rates in the United States. In this paper, a change point was defined as any fundamental or visible shift in interest rates caused by any circumstance. The shift points are then grouped and utilized to anticipate interest rates. When contrasted to the conventional neural network, the neural network model that included change points as one of its input performed better at a 1% significant level. This implies that alternative inputs impacting interest rates can be explored to improve interest rate prediction accuracy.

As per (Hong and Han; 2002), when data mining methods are coupled with previous knowledge, the algorithms are significantly more efficient. To divisional structure news articles as favorable or bad for the economic and financial marketplaces, an expertise approach was utilized as a feedback to a neural network. Commodity prices, money, the monetary policy rate, the unemployment rate, and economic progress are among the other financial and economic factors incorporated in the neural network. As per the report's results, the efficiency of this hybrid system was substantially better than that of a pure neural network.

An artificial neural network (ANN) as well as a non-linear, non - parametric model termed support vector machine (SVM) were tested to anticipate UK interest rate ranges in (Jacovides; 2008). Mostly in history, the researchers questioned the utility of SVM and NN as a substitute for linear and conventional methods. They also stated that by using the term structure information, they could generate strong and extremely accurate interest rate forecasts. When the precision of SVM and NN is compared, the SVM takes the lead.

The Indonesian bank used a backpropagating algorithm to determine interest rates in (Sovia et al.; 2018). Backpropagation of neural networks is one of the learning strategies that has been developed to decrease output error. It is divided into three phases, each of which is repeated until specific criteria are met. In order to forecast interest rates, a variety of factors were utilized in this study, including money supply, JCI, dollar exchange rates, and inflation, among others. Additionally, the intended objective, which is also the rate value, is attained with a backpropagation network design composed of four input levels, two hidden layers, and one output layer. This architecture has an accuracy of 99.73 percent. However, the precision with which interest rates are forecasted is determined by the tolerance threshold used in the back propagation computation procedure.

This subsection gives a brief overview about various basic machine learning models that have been used over time by using backpropagation and error terms are input to neural network and the improvements in performance by using certain variables. The baseline approach that can be induced from this subsection is the use of multiple financial and economic variables as an input to a neural network model.

2.3.2 Hybrid Neural Networks

Another short-term interest rate research (Malliaris and Malliaris; 2007) examined four distinct approaches, including time series, Taylor, neural networks, and econometric. The Taylor rule approach is one of the most often used ways for predicting prospective short-term interest rates. It asserts that any movement in short-term lending rates is correlated with changes in market inflation and potential production. The study separated the range of data into homogenous segments to improve the prediction power of the models and indicated that artificial neural forecasting works better whenever the set of data is split into 3 levels of funding: low, medium, and high. Furthermore, whether interest rates are extremely high or extremely low, neural network models perform best.

In another study (Dua et al.; 2008), interest rate estimation for Indian economy is done using univariate models such as the autoregressive integrated moving average (ARIMA) model and ARIMA models with autoregressive conditional heteroscedasiticity (ARCH) and generalized autoregressive conditional heteroscedasiticity (GARCH) (GARCH). These models were also contrasted to multivariate analysis such as the vector autoregressive model (VAR), vector error correction model (VECM), and Bayesian vector autoregressive models (BVAR). Among the univariate models, the ARCH model is shown to become the most trustworthy for out-of-sample predictions. Because of market volatility, the ARCH and GARCH systems outperformed the ARIMA model. It also indicates that choosing factors for multivariate analysis is an essential component of the research and that multivariate models outperform univariate models in projecting long-term lending rates. The BVAR model, which surpassed all the other simulations, incorporated liquidity, inflation, forward premium, repo-rate, and libor. Nevertheless, the BVAR model itself has disadvantage: it requires the usage of an entity to improve out-of-sample predictions, which makes forecasts outside of the specified period non-optimal. As a consequence, a framework capable of making optimum out-of-sample forecasts with precise input parameters may be studied.

In (Yu and Zhang; 2008), the autoregressive moving average model (ARMA), which is commonly used to predict financial series data for the period, and ANN, which properly represent the nonlinear behavior of time series financial data, were contrasted with a composite ARMA-Elman model. ARMA models are logistic regression that take into account past values, an arbitrary term, and prediction errors, and they presume that the essential information for predicting financial data is stored in the previous data of time series and their prior error terms (Granger and Newbold; 2014). In (Yu and Zhang; 2008), various categories of hybrid ARMA-Elman models were constructed, one using innovations and another with actual data points as intake. When the outputs of all four models were evaluated, the two hybrid systems surpassed the single component and single ARMA models. The hybrid model with input from techniques surpassed some other hybrid system.

Random Forest, SVM, Logistic Regression, Decision Tree, K-nearest neighbor, RNN, and Long Short-Term Memory (LSTM) were used to anticipate the Japanese long-term rate of interest (Suimon et al.; 2019). This study evaluated the forecast accuracy of algorithms that utilized solely Japanese rate of interest market data with algorithms that incorporated the influence of US market data on Japanese lending rates. The outcomes of this study demonstrate that integrating US and Japanese interest rate data can considerably enhance the predictive performance of Japanese long-term lending rates. Modifications in different international interest rate factors, such as foreign banking system monetary policy and foreign financial crises, have been discovered to get a direct influence on the Japanese exchange rate market. When Japanese bank rate data was combined with US bank rate data, every one of the models utilized in (Suimon et al.; 2019) improved significantly.

Utilizing incident sentiment classification as a parameter to the deep neural network improved prediction of Hong Kong interest rates significantly. According to (Yasir et al.; 2020), providing Twitter sentiment classification of six mega-events as feed to a neural network, coupled with regular data on central bank interest rate, leads in a 266 percent drop in inaccuracy. The Neural network surpassed the SVM and linear regression methods.

Concerning interest rate prediction in Montenegro (Pejović and Karadžić; 2020b), the Box-Jenkins technique and autoregressive (AR) models were evaluated. Based on the monthly information on weighted average lending rates of banks, these models were determined to be appropriate for projecting interest rates in Montenegro. The paper also compares the unitary AR model to the multivariate VAR model, indicating that the multivariate VAR model surpasses the AR model in terms of evolution measures such as root mean squared error, forecast error, mean absolute error, and mean squared error. However, the authors suggest that other techniques, such as ANN, ARCH, GARCH, and others, may generate more diverse and trustworthy findings.

The prediction of interest rates using ANN has been improved in (Farahani; 2021) by incorporating novel heuristic approaches such as Moth Flame Optimization (MFO), Time-varying Correlation Particle Swarm Optimization (TVAC-PSO), Chimp Optimization Algorithm (CHOA), and the use of 17 new variables as input to the network such as house price, oil price, gold price, and so on. Various loss functions, including as MSE, RMSE, and estimation errors, were used to evaluate different methods. The application of these innovative metaheuristic algorithms for forecasting long-term interest rates has enhanced computation speed, decreased model complexity, and boosted accuracy. Finally, it has been concluded that the Whale Optimization Algorithm (WOA) outperformed other methods and produced fewer errors.

The hybrid models examined in this part show that including factors other than financial and economic variables can lead to greater accuracy in forecasting interest rates. In addition, many models with modifications and hybrid systems have been created to attain improved performance in forecasting long and short term interest rates.

Brief summary of all the machine learning models analysed in the literature review has been provided in Table 1:

Table 1: Summary details of the related work for forecasting of interest rates

Author(s)	**Objective**	**Data Collection Methods**	**Data Analysis Methods**	**Key Findings**
(Oh and Han; 2000)	Alternative variables to increase accuracy of predictions	Fisher's theory to select variables	RMSE, MAP, MAPE of both models compared	Change points detection and usage gave better results than pure NN at 1% significance level
(Hong and Han; 2002)	Using KBN miner as Congnitive Maps for NN input	News mining from Korean newspapers and financial variables	MAE, MAPE, t-test	HYbrid NN MAE value lowest of 0.527 against Random walk MAE of 1.7 and NN MAE of 0.586
(Malliaris and Malliaris; 2007)	Comparison of 4 methodologies - Taylor, Time Series, Econometric, NN,	Monthly data from Fed Funds	Data division in 5 sets and analysing intercept and coefficient values	Separating data into homogeneous segments increases accuracy and NN outperformed other models
(Dua et al.; 2008)	Univariate and Multivariate models	Variables from Indian economy	Theil's U, e Diebold–Mariano and Pesaran– Timmermann tests	Multivariate models are more accurate for long term forecating
(Sovia et al.; 2018)	Backpropagation neural network	Dollar Exchange Rate Index, the money supply, inflation	MAE, MAPE using Matlab	Precision 99% while using backpropagation
(Suimon et al.; 2019)	Analysing effects of overseas interest rates on Japan's interest rates	U.S.A and Japan's interest rates data	Accuracy	Combining Japan and US rates gave better prediction accuracy
(Pejović and Karadžić; 2020b)	Box-Jenkins,AR and VAR models	Central bank of Montenegro	Autocorrelation, Heteroscedasticity, Residual normality	Weighted average rates used in multivariate VAR model gives more adequate estimates
(Farahani; 2021)	Incorporating novel meta heuristic models with ANN	Financial and economic variables	MSE, RMSE, estimation errors	Novel metaheuristic algorithms improved accuracy, computational speeds

3 Methodology

The technique suggested for this research paper has the primary objective of researching the link between numerous economic factors and assessing the influence of an epidemic on long-term interest rates, in addition to projecting a country's interest rates. Another goal of the suggested technique is to investigate previous links between various economic theories and interest rates. Since banks' liquidity requirements may have expanded during the crisis, it is essential to analyze the sustainability of the link between key financial and economic factors and their influence on interest rates. Knowledge Discovery in Databases (KDD), a knowledge discovery approach, becomes important for assessing these features (Diaz et al.; 2016). Fig. 1 depicts the approach that was developed for this study endeavor.

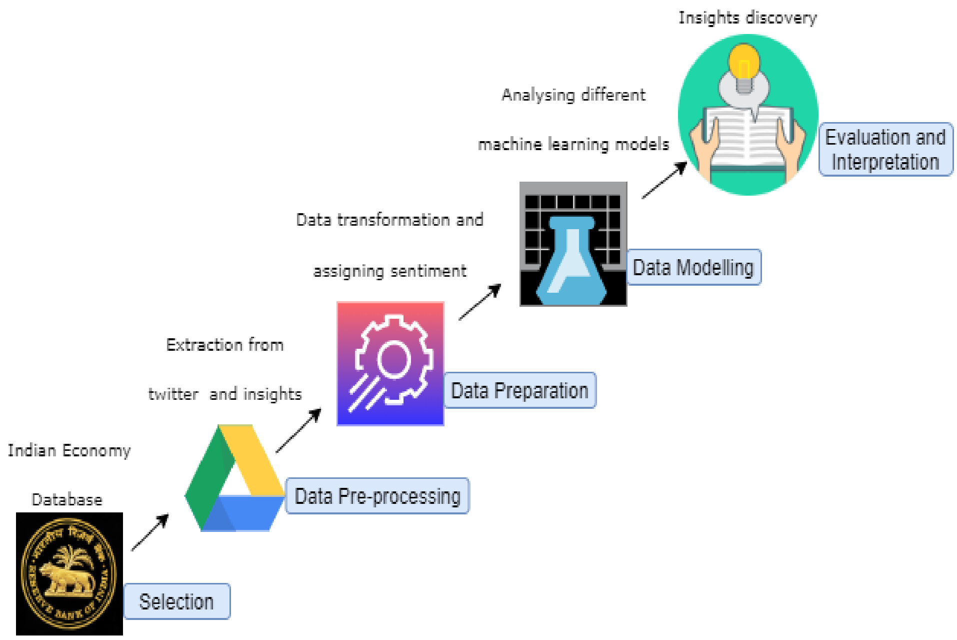


Figure 1: Methodology for forecasting interest rates

3.1 Data Selection

To anticipate interest rates, numerous financial and economic factors must be evaluated and incorporated in the planning method. As a consequence, the Reserve Bank of India (RBI) data repository was selected as the source for this research study on the Indian economy. Statistics on numerous financial and economic indicators, and also interest rates, are accessible on the RBI's data warehouse under the time series publishing databases starting April 2011 to the current period to anticipate India's interest rates. Foreign Currency exchange Deposits, Net Assets, Money Supply Deposit Money, Liquidity Activities, Treasury Notes, and other financial variables were all collected and used for evaluation and also as data to machine learning models. In addition, data from Twitter tweets containing information on various financial events and economic worries was collected to do sentiment analysis. The financial and economics tweets has been extracted in English language and for a period from 2019-20.

3.2 Data Pre-processing

The data from India Economy's data warehouse was analyzed and organized based on the variables included in the csv files. Furthermore, because financial datasets contain a higher percentage of missing and incorrect values, it is essential to keep records of them and effectively alter them by deleting or introducing factors, altering their values, or utilizing the log characteristics of the parameters. This data was verified for missing values or symbols, as the data collection did not contain any values for weeks or even months. Null values in the dataset might cause machine learning models to be incorrectly functional such that all null data points in the data set have been converted on the mean or median basis. It is necessary to examine the data retrieved from twitter before going to the next stage, and clear it since it could include irrelevant or misleading information from the tweets. Twitter data were also examined for null values and tweets were filtered according to their language. After the sentiment analysis the two datasets were combined and finalized as an input into the master learning models, certain specific characters were deleted from the tweets.

3.3 Exploratory Data Analysis

Analyzing exploratory data not only involves identifying the parameters and the data to be analyzed, but also reviewing and analyzing data conceptually and statistically in view of the study aims. It is all part of the theoretical component to find and evaluate proper and relevant characteristics and also to comprehend their behavior in the modification of interest rates. The statistical part of the research involves evaluating how the important variables are distributed and consistent. A correlation matrix comprising some of the most significant aspects, which are shown in fig. 2, has been assessed to analyze the dependence of all financial and economic factors against each other. In addition, various outliers, bias, seasonality, trends etc. have been analyzed for the accurate and relevant information to be provided by the machine learning models.

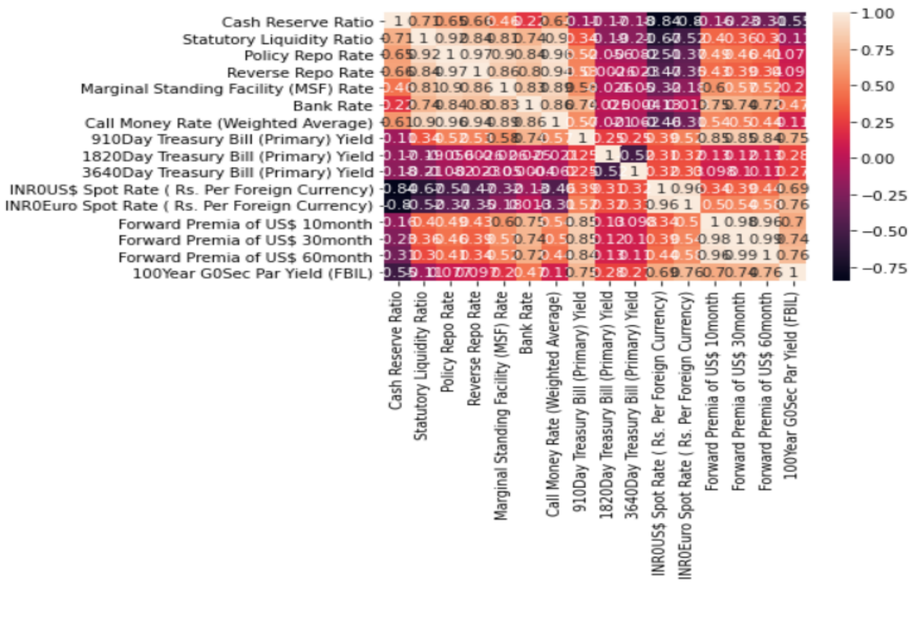


Figure 2: Correlation matrix of financial variables

3.4 Modelling

It is vital to pick the correct Machine Learning model to assess the intended outcomes after pre-processing and analysis of the raw data. For these steps, four distinct models were selected based on the literary analysis of several prediction models for forecasting interest rates. Since it is a time series, data collected from the Indian economy data warehouse publishing series have been assessed and long short term memory model appraised. The approach to a multivariate model has also been adopted, using all financial and economic data as an input to the profound learning model. The novel strategy used to analyze the effect of sentiment in the forecasting of interest rates was the utilization of several financial and economic variables together with the analysis of the sentiment of financial twitter. The analysis was assessed by the construction of a two deep learning model namely: sequential neural model and multilayer perceptron model.

3.5 Knowledge Discovery

The performance of machine learning models, together with the construction of a hybrid neural network, have been evaluated and measured in the various evaluation methods discussed in the evaluation section, to understand the impact of various financial and economic indicators and variables on interest rate values. All models were deployed and inconsistencies were noticed and processed in the models. In addition, the elements that affect mistakes and model accuracy were handled and analyzed carefully.

4 Design Specification

Figure 3 shows the design requirements of this research. In particular, the process design was split into three segments: the database layer, the application layer, and the display layer. The following are the key steps in each tier.

4.1 Database Layer

Because a data-driven strategy was used throughout the research, this is the most essential layer and serves as the source of information for the succeeding layers. The actions taken in this layer are as follows.

1. Obtaining data from the Indian Economy's data warehouse website, which contains information and numbers for different financial and economic factors, as well as interest rates for the last decade.

2. Using the twitterscraper library to extract information from Twitter and save it to a json file.

3. Setting up the GitHub profile to submit updates at regular intervals to avoid data or progress loss.

4. Python is used to process the data so that it may be used in machine learning models.

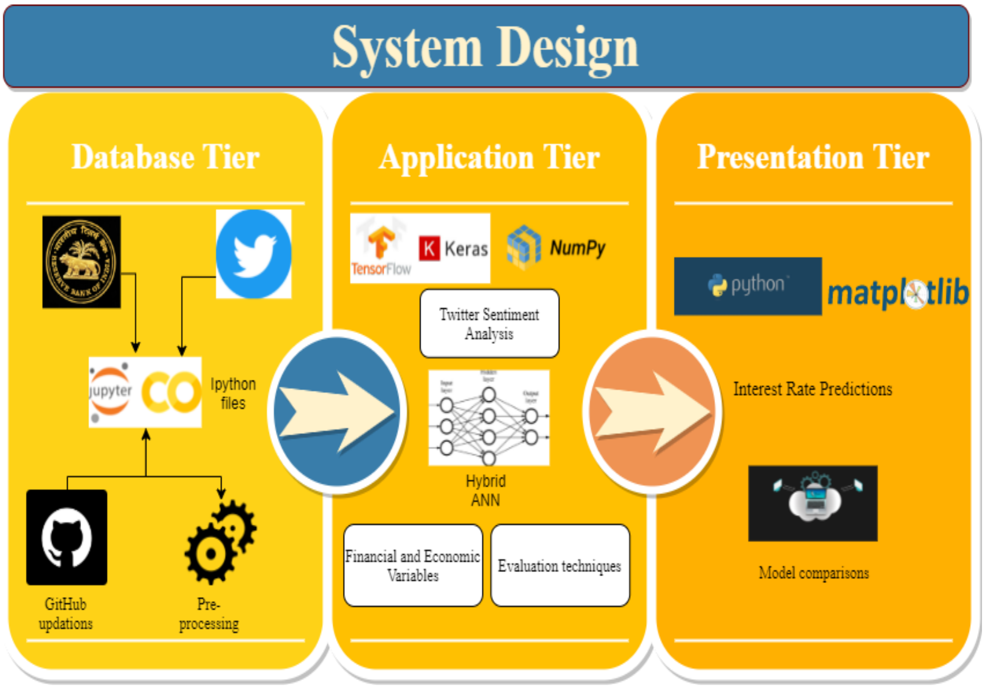


Figure 3: System design followed for the research

4.2 Application Layer

The structure of the hybrid machine learning algorithm has been programmed in this tier. This layer was also used to evaluate various settings in order to get the greatest overall performance of interest rate prediction. The following steps has been performed in this layer.

1. Creating the baseline model as well as the novel machine learning model, the hybrid deep learning model.
2. Separation of datasets into train and test modules.
3. Fine-tuning the algorithms and changing the layers of machine learning algorithms.
4. Using suitable assessment methods to reduce mistakes in interest rate estimate.

4.3 Presentation Layer

After the model produces its output, it is displayed and evaluated in the presentation layer. Steps performed in the presentation layer are as follows.

1. Plotting and visualizing the assessment metrics for all machine learning methods used.
2. A comparison of the baseline and novel machine learning models' outcomes.
3. Summarizing the findings based on the error rates achieved while predicting interest rates.

5 Implementation

5.1 Setting up the environment

The computer utilized for this project is a 64-bit Windows 10 PC with 8 GB of RAM. To scrape Twitter data, the complete program was written in jupyter notebook and Google colab. The research's progress has been regularly posted on GitHub, ensuring that no progress is lost. Because scraping data from Twitter necessitates faster GPUs, twitter scraping through the twitterscraper library was carried out on Google colab. All of the remaining programming stages were written in ipython files using Anaconda's jupyter notebook.

5.2 Data Transformation

At several phases, the data selected for the study was modified to meet the model's criteria. There were two sets of datasets that were used to forecast interest rates.

5.2.1 Financial and Economic Variables

This dataset was obtained from the Indian Economy's data warehouse and included a time series dataset of different financial and economic ratios and values such as cash reserve ratio, bank rates, credit deposit ratio, and 23 other similar variables from 2011 to 2021. The dataset is updated weekly and consists of 550 rows and 26 columns, as well as the goal variable, the policy repo rate. Following an analysis of the correlation matrix for all variables, certain variables were removed from the dataset due to their uncorrelated character in the prediction of interest rates. In addition, the dataset has been checked for missing values. Because the values were determined to be missing not at random (MNAR), the KNN imputer was used to fill in the gaps. The KNN imputer finds missing data using the Euclidean technique and the k-nearest neighbors model to fill in the blanks. It computes the mean value of the parameters' n nearest neighbors. To verify trends, all variables have been represented as time series data, and fig. 4 displays an interactive representation of the target variable, the policy repo rate.

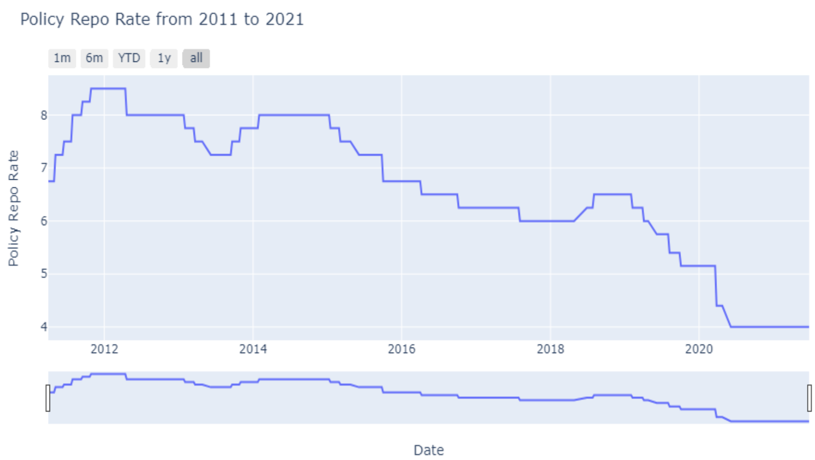


Figure 4: Policy Repo Rate

5.2.2 Twitter Data

Using the Twint technology, tweets about India's economic policies were removed from Twitter. Twint is a Python-based Twitter scraping program that uses Twitter's search operations to collect tweets based on specific searches such as users, pre-defined topics, hashtags, and so on. Popular tweets between 2011 and 2021 were scraped and saved in a CSV file after successfully setting the environment and loading libraries for utilizing twint. The Python TextBlob package was used to assign feelings to tweets, which were then transformed to -1, 0, and 1 for negative, neutral, and positive emotions.

5.2.3 Merging data for novel approach

One of the most difficult problems in evaluating the innovative technique has been combining data from all financial and economic factors with feelings from tweets. The complete collection has been split into weekly data by extracting weeks from dates as well as their years. Furthermore, based on the weekly emotions from tweets, both datasets were combined using a left join. Figure 5 depicts some missing values in the combined datasets that were handled by the KNN imputer.

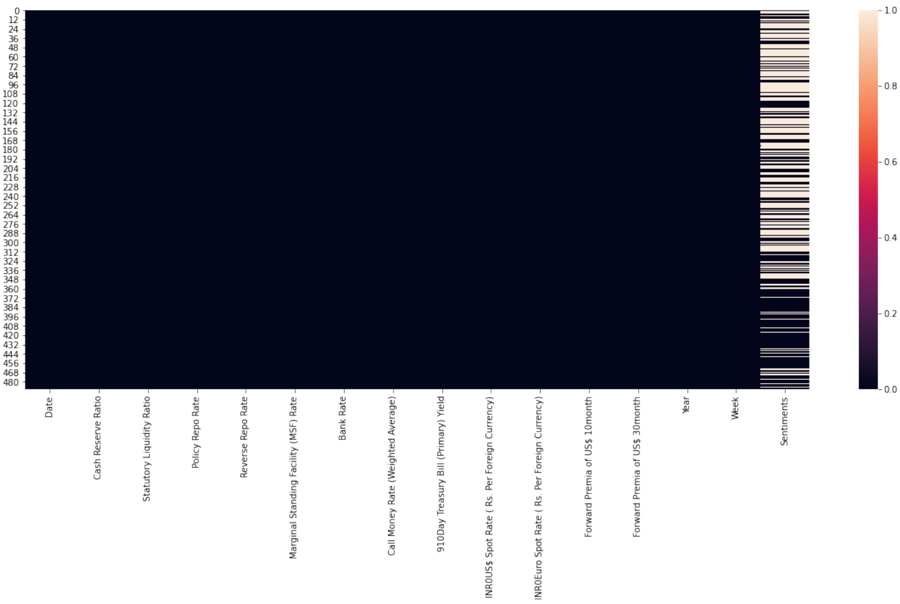


Figure 5: Missing values in the merged dataset

5.3 Data Modelling

For conducting this research model selection has been done considering the forecasting of interest rates as a regression time series problem. Based on the literature review the baseline approach has been selected as using multiple variables with VAR and sequential neural network models. Therefore a total of 6 models have been built and analyzed which are VAR multivariate model, LSTM model, MLP model (baseline and novel), and a sequential deep learning model (baseline and novel). The pre-processed and transformed data has been used as an input to these models. Fig. 6 depicts the process flow diagram about the implementation approach for this research.

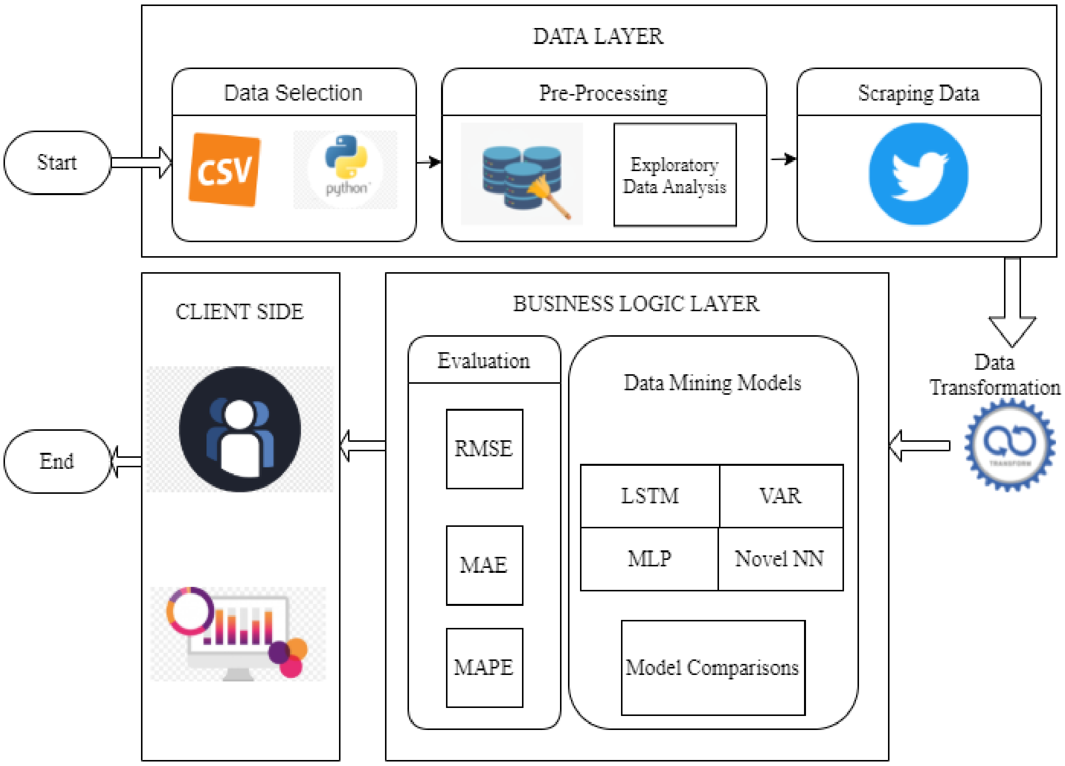


Figure 6: Process flow diagram

5.3.1 Vector Autoregressive multivariate model

The Vector Autoregressive model is a multivariate algorithm that is used when multiple time series affect each other. It is one of the most important modeling techniques used to forecast financial time series data because each of the variables is evaluated based on its past values. Since multiple equations are used so it factors the bi-directional nature of the variables.

$$k_t = A_0 + A_1 k_{t-1} + A_2 k_{t-2} + A_2 k_{t-2} + \cdots + A_n k_{t-n} + e_t \tag{1}$$

where $k_n = (k_{1t}, k_{2t}, \ldots, m_{Kt})$ for $k = 1, \ldots, K$ time series. The key assumptions are in the VAR(p)

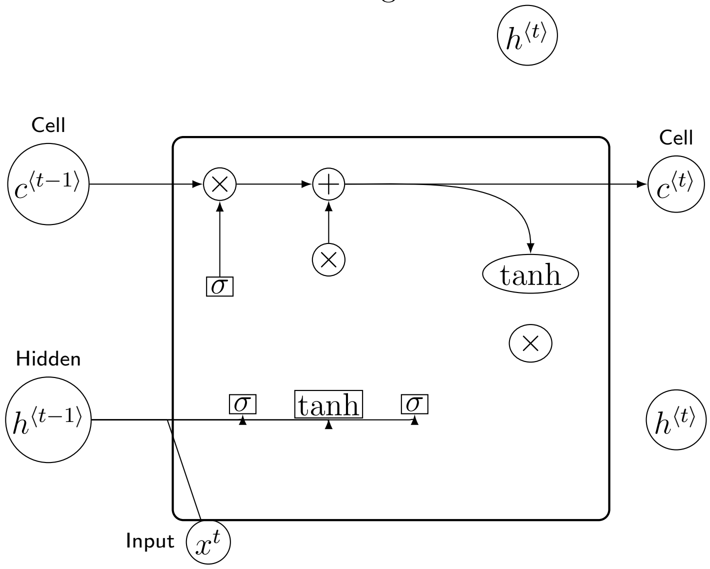
$$\Delta k_t = \Delta\Gamma 1 k_{t-1} + \Delta\Gamma 2 k_{t-2} + \cdots + \Delta\Gamma n k_{t-n} \tag{2}$$

Where $\gamma_i = -(I - A_1 - \cdots - A_i)$, $i = 2, \ldots, p-1$ $\Pi = -(I - A_1 - \cdots - A_p)$ is a N-dimensional time series, A_0 is the intercept term, e_t is white noise.

The MAPE value when this model has been evaluated on the financial and economic variable dataset is 2.43% which is considered excellent in terms of forecasting.

5.3.2 Long short term memory model

The Lstm is a specific form of neural network (RNN) utilized for the learning of long-term interconnections. Basic components of an LSTM network are units or logic gates. There seem to be 2 distinct states that are the control unit and the covered state. The chunks of memory have doors that recall and modify things. LSTM has three gates namely the output gate, gate inputs and forget gate. The LSTM is utilized as a continuous layer load or sequence model throughout this study. Following is the representation of how a LSTM network functions with gates at each level of data manipulation.



The MAPE value of 2.88% has been achieved when this model has been applied on the target variable i.e. the policy repo rate.

5.3.3 Sequential deep learning model

Sequential neural networks are machine learning models that cope with data sequences. These models are typically used to forecast time series or regression issues, and any of the input or output might be a sequence. It enables the construction of models using a layer-by-layer technique, with each layer having its own set of weights. These neural networks model datasets using a collection of Python functions, such as the add function, which is used to create several layers. This neural network has been employed in this study by applying it to both the baseline and novel approaches of utilizing sentiments from tweets.

1. Baseline approach: The dataset has been separated into test and train subsets in order to feed the dataset with various financial and economic factors as input to the sequential neural network. The StandardScaler function has been used to scale both the training and testing sets because it removes the mean values from the variables and assesses the unit variance of each variable. This has been done to avoid the effect of any big value range on the distance computation between features in the neural network. During model construction, the SGD optimizer has been also employed since it aids in loss reduction by controlling the neural network's learning rate and weights. To assess the model, the sequential neural network has been built with the following evaluation measures as loss functions: mean squared error, mean absolute error, and mean absolute error percentage. Fig. 7 represents the evaluations derived from the output of the baseline approach sequential neural network.

```
The Root Mean Square Error for Baseline approach sequential NN model is 0.04
The Mean Absolute Error for Baseline approach sequential NN model is 0.04
The Mean Absolute Percentage Error for Baseline sequential approach NN model is 5.24%
```

Figure 7: Evaluations for baseline approach on sequential neural network

2. Novel approach: The basic structure of the sequential neural network has been retained the same as in the baseline method, but the neural network's input data now includes sentiment from tweets as well as different financial and economic indicators. The merged dataset has been trained, and fig. 8 depicts the neural network's output evaluations.

```
The Root Mean Square Error for Novel approach sequential NN model is 0.07
The Mean Absolute Error for Novel approach sequential NN model is 0.04
The Mean Absolute Percentage Error for Novel sequential approach NN model is 4.80%
```

Figure 8: Evaluations for novel approach on sequential neural network

5.3.4 Multilayer perceptron model

Deep learning methods with single or many layers of neurons are based on multilayer perceptrons models. The modified data is supplied into the MLP's input layer, and hyperparameter optimization is performed to determine the appropriate number of hidden layers, since these hidden layers offer abstraction levels for predictions. The output layer's output is examined. They are frequently used for classification issues, but they are also very good at predicting regression problems since the target variable in a regression problem is frequently reliant on a collection of other predictor factors.

1. Baseline approach: The financial and economic variable dataset has been divided into training and testing sets and again the StandardScaler function has been used to transform the training and testing sets. A relu activation with hidden layer size as (64,32) has been initialized. Further, the predictions from the MLP model have been evaluated based on mean square error, mean absolute error, and mean absolute error percentage and have been depicted in fig. 9.

```
The Root Mean Square Error for baseline MLP model is 0.18
The Mean Absolute Error for baseline MLP model is 0.13
The Mean Absolute Percentage Error for baseline MLP model is 2.09%
```

Figure 9: Evaluations for baseline approach on MLP neural network

2. Novel approach: For this approach, the input data used has been merged with sentiments from tweets and the best evaluations have been noted to derive with hidden layer as (128,64,32). The results of the evaluations on predictions from this model have been depicted in fig. 10.

```
The Root Mean Square Error is for novel MLP model is 0.16
The Mean Absolute Error is for novel MLP model is 0.11
The Mean Absolute Percentage Error is for novel MLP model is 1.76%
```

Figure 10: Evaluations for novel approach on MLP neural network

6 Evaluation and Results

Verification of results from machine learning techniques to assess achievement of research objectives is an essential step. Because the suggested challenge of predicting interest rates may be analyzed as a regression problem, the evaluation metrics used to evaluate a regression problem were employed in this study. For this study endeavor, the following assessment metrics were employed.

1. Root Mean Square Error (RMSE) : The deviation of error terms or loss in forecast from the main regression line is denoted as RMSE. It is commonly referred to as the residual spread, and it represents the concentration of data points surrounding the line of best fit. It is evident from equation 3 that the RMSE value provides an indication of how much our forecasts deviate from the original data.
$$RMSE = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}\Big(forecast(x) - actual(x)\Big)^2} \tag{3}$$
2. Mean Absolute Error (MAE): MAE is a method for calculating the average degree of errors in a collection of predicted values without taking into account their distribution. All mistakes are given identical weights, but because interest rates might be negative, this assessment parameter is significant. Equation 4 represents the formula for the calculation of mean absolute error.
$$MAE = \frac{1}{n}\Sigma_{i=1}^{n}|forecast(x) - actual(x)| \tag{4}$$
3. Mean Absolute Percentage Error: MAPE, also known as mean absolute percentage deviation (MAPD), is an assessment of a projected value's prediction accuracy. Equation 5 shows that the MAPE is a measure of the forecast error and thus helps in identifying fittness of a model. The lower the value of MAPE the more fit and accurate a model is.
$$MAPE = \frac{1}{n}\Sigma_{i=1}^{n}|\frac{1}{F}A - F| * 100 \tag{5}$$
Where A- actual value, F- forecasted value, and n- number of observations

6.1 Results

Based on the evaluation metrics described above the results of all the models applied have been defined in this section. It is notable that the lower values of RMSE, MAE, and MAPE are considered to excellent in terms of forecasting by a model.

6.1.1 Vector autoregressive model

Fig. 11 displays the forecasted and actual value of the repo rate over a period of 6 months. It is evident that the forecasted repo rate is not very far from the actual repo rate and the model produced a mape of 2.43%.

6.2 Long shot term memory model

Fig. 12 displays the forecasted and actual repo rates. It has been noted that the forecasted rates have a constant value because the rates have not changed their value in the past 1.5 years and since LSTM model learns from its past value therefore the forecaste values are the same with a mape value of 3.12%.

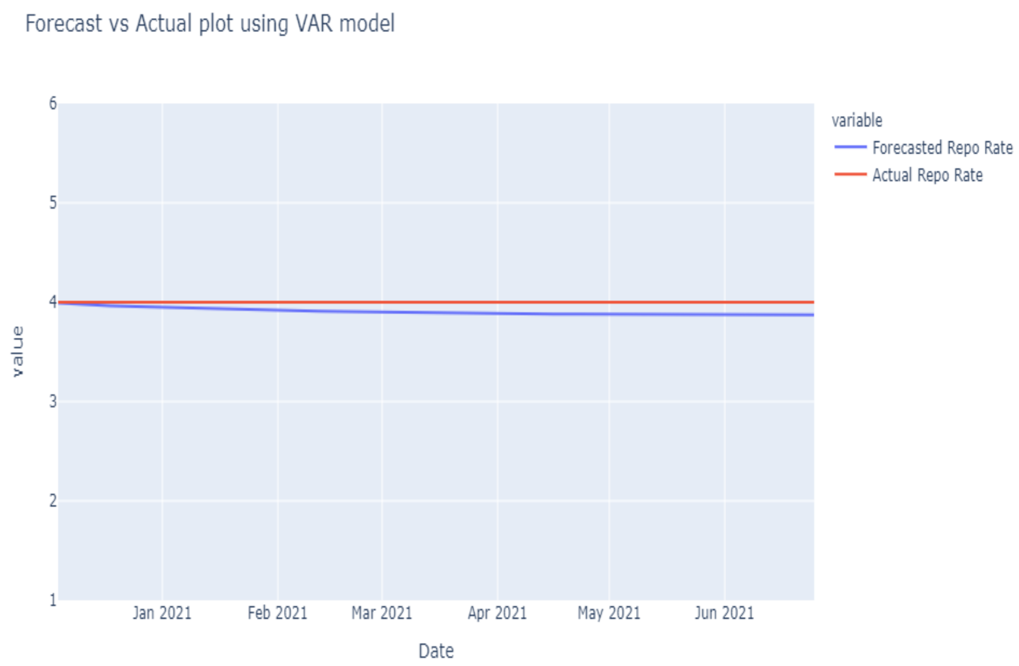


Figure 11: Predictions vs Actual Repo Rate VAR model output

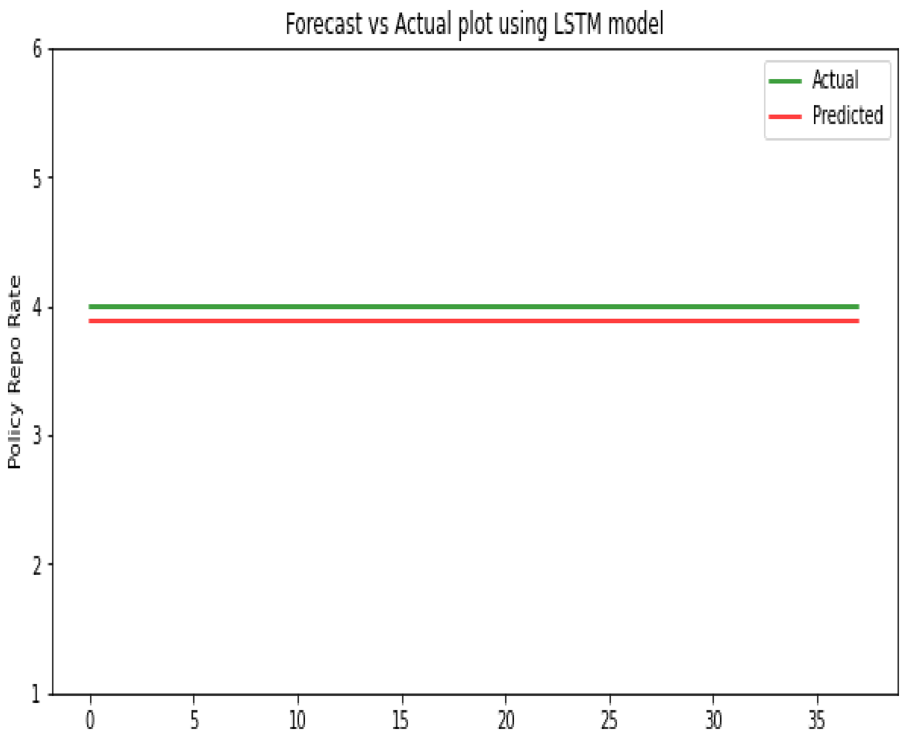


Figure 12: Predictions vs Actual Repo Rate LSTM model output

6.3 Sequential neural network

6.3.1 Baseline approach

Fig. 13 shows the plot of loss function generated during modeling the baseline approach for the sequential neural network. The value of rmse for this approach is 0.040.

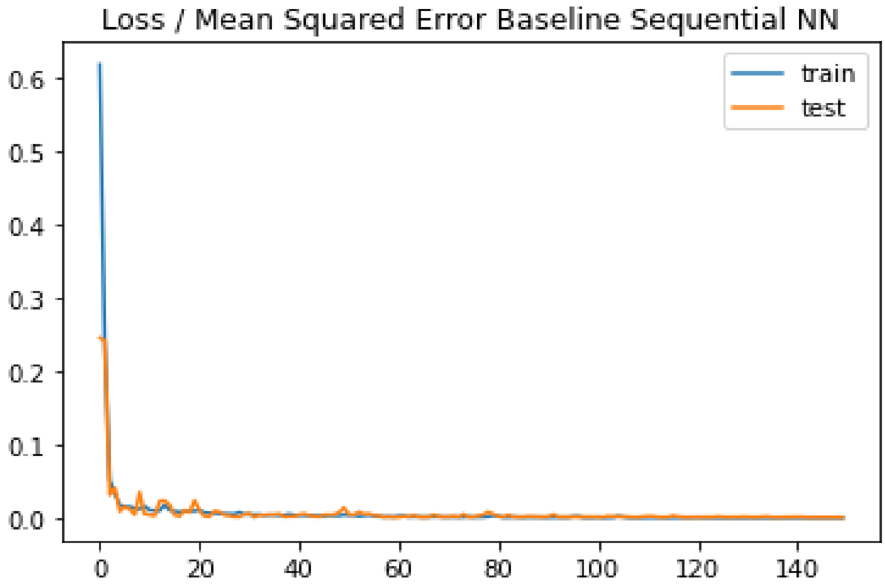


Figure 13: Loss/Mse for baseline approach sequential NN

6.3.2 Novel Approach

Fig. 14 shows the plot of loss function generated during modeling the novel approach for the sequential neural network. The value of rmse for this approach is 0.070.

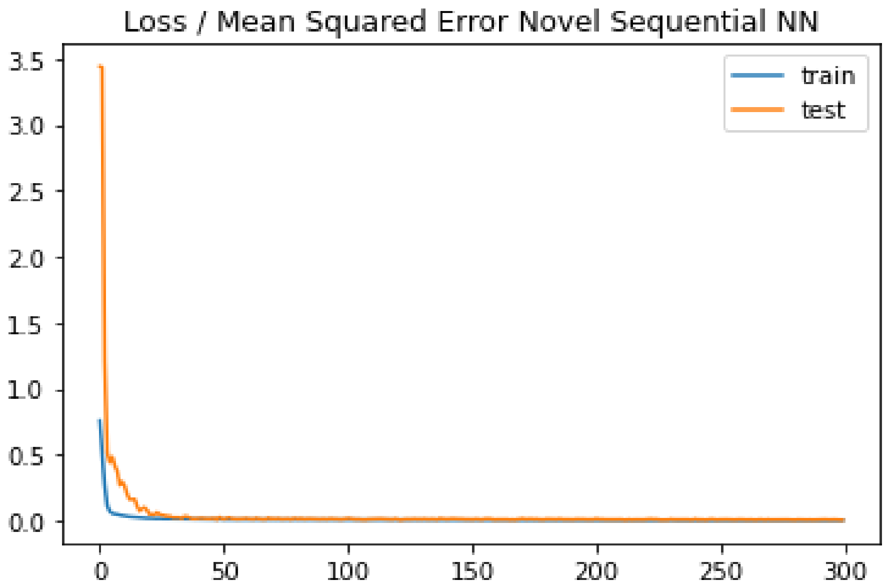


Figure 14: Loss/Mse for novel approach sequential NN

6.4 MLP neural network

6.4.1 Baseline approach

Fig. 15 represents the forecasted and actual values of the repo rates. As evident from the graph as well as the mape value this model has proved to be quite efficient as the mape value is 2.09%.

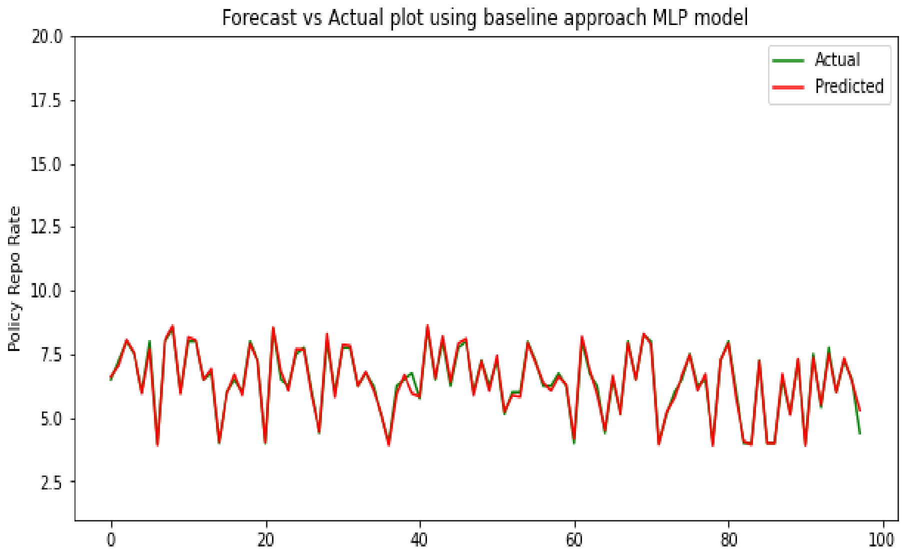


Figure 15: Predictions vs Actual Repo Rate Baseline MLP model output

6.4.2 Novel Approach

Fig. 16 represents the the forecasted and actual values of the repo rates. The novel approch has proved to be even more significant than the baseline approach as the mape value is 1.76%.

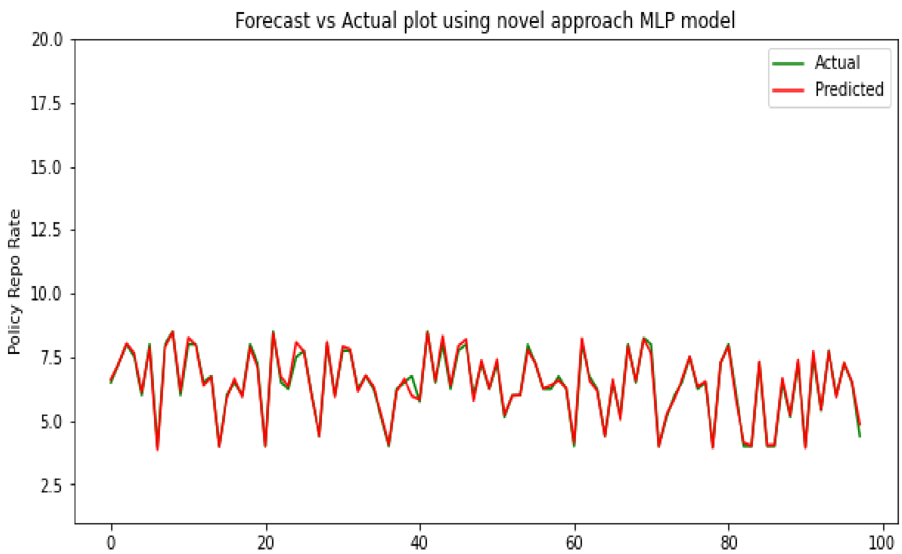


Figure 16: Predictions vs Actual Repo Rate Novel MLP model output

Fig. 17 displays a table with comparison of evaluation metrics applied to all the models. Both the novel models have proved to be more significant from their baseline approaches as the rmse, mae, and mape values are lesser for the novel approach. After comparing the models it has been noted that the use of sentiment analysis from twitter regarding repo rates have proved to increase the forecasting accuracy of the sequential as well as the MLP neural network models. However, the best and most significant forecasts have been done by the MLP neural network by using all the financial and economic variables as well as the sentiment analysis from twitter data as input.

	VAR	LSTM	Baseline Sequential NN	Novel Sequential NN	Baseline MLP	Novel MLP
RMSE	0.100	0.120	0.040	0.070	0.180	0.160
MAE	0.190	0.120	0.040	0.040	0.130	0.110
MAPE	2.43%	3.12%	5.24%	4.80%	2.09%	1.76%

Figure 17: Comparison of all the models for forecasting of interest rates

7 Conclusion and Future Work

Forecasting interest rates is a crucial step in anticipating future actions of any government policy or developments in any financial ventures. The goal of this study was to identify all of the important elements that influence interest rate variations and can help enhance interest rate forecasting accuracy. The use of sentiment analysis from tweets as an input to machine learning systems has proven to improve interest rate forecasting accuracy. Scrapping tweets from Twitter and combining the sentiment of tweets with other financial and economic variables has proven to reduce forecasting inaccuracy, demonstrating that factors such as public opinion about the economy play a major part in interest rate swings. The novel sequential and MLP models performed better than their state-of-the-art approaches and when all the machine learning models employed in this study were compared, the novel MLP deep learning model produced the lowest mean absolute percentage error of 1.76 percent and beat all other models in producing most relaible results.

The future work includes finding more relevant features that can further enhance the prediction power of the machine learning models such as policy changes in other countries or estimates about various financial variables from certain financial and economic reports. Also, the behavior/sentiment of the officials responsible for fluctuations in interest rates based on their past actions could be analyzed and used as a factor to predict interest rates.

References

Bernanke, B. S., Boivin, J. and Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (favar) approach, *The Quarterly journal of economics* **120**(1): 387–422.

Bouqata, B., Bensaid, A. and Palliam, R. (1999). Recurrent artificial neural networks for forecasting of forward interest rates, *IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No.99CH36339)*, Vol. 6, pp. 3894–3897 vol.6.

Chan, K. C., Karolyi, G. A., Longstaff, F. A. and Sanders, A. B. (1992). An empirical comparison of alternative models of the short-term interest rate, *JOURNAL OF FINANCE* **47**(3): 1209–1227.

Cox, J. C., Ingersoll Jr, J. E. and Ross, S. A. (2005). A theory of the term structure of interest rates, *Theory of valuation*, World Scientific, pp. 129–164.

Diaz, D., Theodoulidis, B. and Dupouy, C. (2016). Modelling and forecasting interest rates during stages of the economic cycle: A knowledge-discovery approach, *Expert Systems with Applications* **44**: 245–264.
URL: *https://www.sciencedirect.com/science/article/pii/S0957417415006235*

Dua, P., Raje, N. and Sahoo, S. (2008). Forecasting interest rates in india, *Margin: The Journal of Applied Economic Research* **2**(1): 1–41.
URL: *https://doi.org/10.1177/097380100700200101*

Enke, D. and Thawornwong, S. (2005). The use of data mining and neural networks for forecasting stock market returns, *Expert Systems with applications* **29**(4): 927–940.

Estrella, A. and Mishkin, F. S. (1997). The predictive power of the term structure of interest rates in europe and the united states: Implications for the european central bank, *European economic review* **41**(7): 1375–1401.

Farahani, M. (2021). Prediction of interest rate using artificial neural network and novel meta-heuristic algorithms, *Iranian Journal of Accounting, Auditing & Finance* .
URL: *https://ijaaf.um.ac.ir/article_39902_3bf6e844cbd9c333590e5d8aac1fa689.pdf*

Gallant, A. R. and Tauchen, G. (2001). Emm: A program for efficient method of moments estimation, *User's Guide, Department of Economics, University of North Carolina* .

Gordon, A., Sharpe, W. and Bailey, J. (2003). Fundamentals of investments/fundamentos de inversiones.

Granger, C. W. J. and Newbold, P. (2014). *Forecasting economic time series*, Academic Press.

Harrison, R., Nikolov, K., Quinn, M., Ramsay, G., Scott, A. and Thomas, R. (2005). *The Bank of England quarterly model*, Bank of England London.

Hicks, J. R. et al. (1975). Value and capital: An inquiry into some fundamental principles of economic theory, *OUP Catalogue* .

Hong, T. and Han, I. (2002). Knowledge-based data mining of news information on the internet using cognitive maps and neural networks, *Expert Systems with Applications* **23**(1): 1–8.
URL: *https://www.sciencedirect.com/science/article/pii/S0957417402000222*

Jacovides, A. (2008). *Forecasting Interest Rates from the Term Structure: Support Vector Machines Vs Neural Networks*, PhD thesis, Citeseer.

Jie, M., Cuicui, W. and Manying, B. (2008). Modeling on interest rate behavior of rmb: Comparative research based on vasicek, cir and ckls model, *2008 IEEE International Conference on Automation and Logistics*, pp. 2941–2945.

Ju, Y., Kim, C. and Shim, J. (1997). Genetic-based fuzzy models: interest rate forecasting problem, *Computers & industrial engineering* **33**(3-4): 561–564.

Long-Zhen Fan and Lan-Jun Lao (2004). An empirical comparison of alternative models of short interest rate with repo rate in the inter-bank market of china, *Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.04EX826)*, Vol. 5, pp. 2736–2742 vol.5.

Ma Jie, Wang Cuicui and Bai Manying (2008). Modeling on interest rate behavior of rmb: Comparative research based on vasicek, cir and ckls model, *2008 IEEE International Conference on Automation and Logistics*, pp. 2941–2945.

Malliaris, A. G. and Malliaris, M. (2007). Modeling short term interest rates: A comparison of methodologies, *2007 International Joint Conference on Neural Networks*, pp. 732–736.

McCulloch, J. H. (1971). Measuring the term structure of interest rates, *The Journal of Business* **44**(1): 19–31.

Merton, R. C. (1973). Theory of rational option pricing, *The Bell Journal of economics and management science* pp. 141–183.

Oh, K. J. and Han, I. (2000). Using change-point detection to support artificial neural networks for interest rates forecasting, *Expert Systems with Applications* **19**(2): 105–115.
URL: *https://www.sciencedirect.com/science/article/pii/S0957417400000257*

Pejović, B. and Karadžić, V. (2020a). Econometric modeling and forecasting of interest rates in montenegro, **53**(1): 72–83.
URL: *https://www.library.ien.bg.ac.rs/index.php/ea/article/view/1216*

Pejović, B. and Karadžić, V. (2020b). Econometric modeling and forecasting of interest rates in montenegro, *Economic Analysis* pp. 72–83.
URL: *https://www.ceeol.com/search/article-detail?id=895932*

Sovia, R., Yanto, M., Gema, R. L. and Fernando, R. (2018). Bank indonesia interest rate prediction and forecast with backpropagation neural network, *2018 International Conference on Information Technology Systems and Innovation (ICITSI)*, pp. 429–435.

Suimon, Y., Sakaji, H., Shimada, T., Izumi, K. and Matsushima, H. (2019). Japanese long-term interest rate forecast considering the connection between the japanese and us yield curve, *2019 IEEE Conference on Computational Intelligence for Financial Engineering Economics (CIFEr)*, pp. 1–7.

Vasicek, O. (1977a). An equilibrium characterization of the term structure, *Journal of Financial Economics* **5**(2): 177–188.
URL: *https://www.sciencedirect.com/science/article/pii/0304405X77900162*

Vasicek, O. (1977b). An equilibrium characterization of the term structure, *Journal of financial economics* **5**(2): 177–188.

Vela, D. (2013). Forecasting latin-american yield curves: An artificial neural network approach, *Borradores de Economics; No. 761* .

Yasir, M., Afzal, S., Latif, K., Chaudhary, G. M., Malik, N. Y., Shahzad, F. and Song, O.-y. (2020). An efficient deep learning based model to predict interest rate using twitter sentiment, *Sustainability* **12**(4): 1660.

Yellen, J. L., Jianping, Z., Peiying, G., Lu, L. and Xiangyun, J. (2015). Normalizing monetary policy-prospects and perspectives: A speech by janet l. yellen on the conference of frb of san francisco.

Yu, X. and Zhang, J. (2008). A comparison of hybrid arma-elman models with single models for forecasting interest rates, *2008 Second International Symposium on Intelligent Information Technology Application*, Vol. 2, pp. 985–989.

Zurada, J. M. (1992). *Introduction to artificial neural systems*, Vol. 8, West St. Paul.