AN APPLIED APPROACH TO FOREX MARKET PREDICTION USING MACHINE LEARNING

CHENG YUN

UNIVERSITI TEKNOLOGI MALAYSIA

UNIVERSITI TEKNOLOGI MALAYSIA
DECLARATION OF Choose an item.

Author's full name	:	CHENG YUN			
Student's Matric No.	:	MCS241014	Academic Session	:	2024/2025
Date of Birth	:	15/01/1996	UTM Email	:	chengyun@graduate.utm.my
Choose an item. Title	:	AN APPLIED APPROACH TO FOREX MARKET PREDICTION USING MACHINE LEARNING			

I declare that this thesis is classified as:

☒ **OPEN ACCESS** — I agree that my report to be published as a hard copy or made available through online open access.

☐ **RESTRICTED** — Contains restricted information as specified by the organization/institution where research was done. *(The library will block access for up to three (3) years)*

☐ **CONFIDENTIAL** — Contains confidential information as specified in the Official Secret Act 1972)

(If none of the options are selected, the first option will be chosen by default)

I acknowledged the intellectual property in the Choose an item. belongs to Universiti Teknologi Malaysia, and I agree to allow this to be placed in the library under the following terms :

1. This is the property of Universiti Teknologi Malaysia
2. The Library of Universiti Teknologi Malaysia has the right to make copies for the purpose of only.
3. The Library of Universiti Teknologi Malaysia is allowed to make copies of this Choose an item. for academic exchange.

Signature of Student:

Signature :

Full Name: UMMI FARIHAH BINTI ABD WAHID
Date : 24 JULY 2024

Approved by Supervisor(s)

Signature of Supervisor I:	Signature of Supervisor II
Full Name of Supervisor I PROF. MADYA DR MOHD SHAHIZAN OTHMAN	Full Name of Supervisor II
Date :	Date :

NOTES : If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentiality or restriction

This letter should be written by a supervisor and addressed to Perpustakaan UTM. A copy of this letter should be attached to the thesis.

Date:

Librarian

Jabatan Perpustakaan UTM,

Universiti Teknologi Malaysia,

Johor Bahru, Johor

Sir,

CLASSIFICATION OF THESIS AS RESTRICTED/CONFIDENTIAL

TITLE: MODELLING THE EFFECTS OF OIL PRICE AND STOCK MARKET PRICE ON GOLD PRICE USING LONG SHORT-TERM MEMORY (LSTM) AND VECTOR AUTOREGRESSION (VAR) MODEL

AUTHOR'S FULL NAME:UMMI FARIHAH BINTI ABD WAHID

Please be informed that the above-mentioned thesis titled _____should be classified as RESTRICTED/CONFIDENTIAL for a period of three (3) years from the date of this letter. The reasons for this classification are

(i)

(ii)

(iii)

Thank you.

Yours sincerely,

SIGNATURE:

NAME:

ADDRESS OF SUPERVISOR: PROF. MADYA DR. SHAHIZAN OTHMAN

"I hereby declare that I have read this project report and in my opinion this project report is sufficient in term of scope and quality for the award of the degree of Master of Science (Data Science)"

Signature	:	_____
Name of Supervisor I	:	PROF. MADYA DR. MOHD. SHAHIZAN OTHMAN
Date	:	17 JANUARY 2025

Signature	:	_____
Name of Supervisor II	:	
Date	:	

Signature	:	_____
Name of Supervisor III	:	
Date	:	

Declaration of Cooperation

This is to confirm that this research has been conducted through a collaboration Ummi Farihah binti Abd Wahid and University Technologi Malaysia (UTM)

Certified by:

Signature :
Name :
Position :
Official Stamp
Date

* This section is to be filled up for theses with industrial collaboration

Pengesahan Peperiksaan

Tesis ini telah diperiksa dan diakui oleh:

Nama dan Alamat Pemeriksa Luar **:**

Nama dan Alamat Pemeriksa Dalam **:**

Nama Penyelia Lain (jika ada) **:**

Disahkan oleh Timbalan Pendaftar di Fakulti:

Tandatangan :

Nama :

Tarikh :

AN APPLIED APPROACH TO FOREX MARKET PREDICTION USING MACHINE LEARNING

CHENG YUN

A project report submitted in partial fulfilment of the
requirements for the award of the degree of
Master of Science (Data Science)

Faculty of Computing
Universiti Teknologi Malaysia

JANUARY 2025

DECLARATION

I declare that this project report entitled *"Modelling The Effects of Oil Price and Stock Market Price on Gold Price Using Long Short-Term Memory (LSTM) and Vector Autoregression (VAR) Model"* is the result of my own research except as cited in the references. The project report has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature	:	.....
Name	:	UMMI FARIHAH BINTI ABD WAHID
Date	:	17 JANUARY 2025

ACKNOWLEDGEMENT

In preparing this thesis, I was in contact with many people, researchers, academicians, and practitioners. They have contributed towards my understanding and thoughts. In particular, I wish to express my sincere appreciation to my main thesis supervisor, Professor Dr. Mohd Shariff Nabi Baksh, for encouragement, guidance, critics and friendship. I am also very thankful to my co-supervisor Professor Dr Awaluddin Mohd Shaharoun and Associate Professor Dr. Hishamuddin Jamaluddin for their guidance, advices and motivation. Without their continued support and interest, this thesis would not have been the same as presented here.

I am also indebted to Universiti Teknologi Malaysia (UTM) for funding my Ph.D study. Librarians at UTM, Cardiff University of Wales and the National University of Singapore also deserve special thanks for their assistance in supplying the relevant literatures.

My fellow postgraduate student should also be recognised for their support. My sincere appreciation also extends to all my colleagues and others who have provided assistance at various occasions. Their views and tips are useful indeed. Unfortunately, it is not possible to list all of them in this limited space. I am grateful to all my family member.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

FOREX	-	Foreign exchange rate market
LSTM	-	Long Short-Term Memory
COVID-19	-	Coronavirus disease
GDP	-	Gross Domestic Product
CPI	-	Consumer price index
VAR	-	Vector Autoregressive Model
VECM	-	Vector Error Correction Model

CHAPTER 1

INTRODUCTION

1.1 Introduction

In the current world,there are more and more transnational corporations which pay much attention into the FOREX market. Predicting future foreign exchange rates is significant for most transnational corporations and even for domestic central banks,which means enterprises can save massive money on the currency risk and central banks can prepare for future financial risks in advance. Because the fluctuations of the exchange rate can significantly have a great effect on international trade, investment, and domestic financial market stability,the authorities from different countries and lots of transnational enterprises are warned to research good methods analysing the FOREX market . Accurately forecasting future values of the FOREX is necessary for stakeholders to develop effective monetary policies, hedging strategies, and financial planning.

In fact,foreign exchange rates are influenced by a variety of factors, including inflation, interest rates, balance of payments, and other relevant risks. Interest rate plays a key role in the movement of the FOREX market because it directly affects the direction of investor returns and capital flows. Inflation, represented by the Consumer Price Index (CPI), plays a secondary role, with its impact on FOREX becoming noticeable when CPI significantly rises . In addition, the 10 Years bond usually can be confirmed as the future interest value by the financial market.

With the analysis of historical data on interest rates, CPI, and 10-Year bond,it can develop an effective model to address this puzzle,which can forecast future FOREX trends, suggesting stakeholders in making valuable and data-driving decisions. This study use these factors to develop predictive insights, highlighting the relationship between macroeconomic indicators and foreign exchange rate movements.

.

1.2 Problem Background

The frequent fluctuations of foreign exchange lead to substantial significant financial challenges for global businesses and economies. According to the 2022 financial reports of Google and Apple, they lost approximately $1.511 billion USD due to dramatical exchange rate movements. And Apple reported losses of $1.836 billion due to negative foreign exchange movements. These figures emphasize the importance of FOREX on transnational corporations, even minimal oscillation can transform the loss of billions of dollars.

Meantime in this aspect of the country , the consequences of FOREX instability also displays a tragic picture. For an obvious example, because of the rapid devaluation of the Argentine Peso against the US Dollar, Argentina has already aggravated economic and social life. The poverty rate in Argentina skyrocketed to over 52.9% in the first half of 2024, compared to 25.7% in 2018, which means 25 million people suffer from the lack of food, medical service ,education and so on, according to the UCA Social Debt Observatory. This instance showcases that it is important for a country to develop effective tools to predict and mitigate the effects of currency fluctuations on both macroeconomic stability and social well-being.

1.3 Problem Statement

Normally, many people face the problem of lacking valuable financial knowledge, rather than the Forex market. A useful and efficient data predicting models benefit their interest.

International businesses face Forex risk in nearly every transaction, which can decrease interest and increase costs if exchange rates shift unstably. This

uncertainty leads companies to rely on hedging strategies, which can be costly and complex. Developing improved methods for managing currency risk, especially for small and medium-sized enterprises , is a critical issue in maintaining global trade efficiency.

1.4 Research Questions

The proposed project aims to achieve the following objectives:

I. How does the interest rate impact the pairs USD/CNY exchange rate from 2014 to 2024?

II. Which variables have played a significant role in the movement of FOREX?

III. How does the LSTM model forecast the future exchange rate of pairs USD/CNY?

1.5 Research Objectives

The objectives of the research are:

(a) To analyse the trend of interest rate,exchange rate of USD/CNY,CPI and 10-year bond from 2014 to 2024.

(b) To predict the future exchange rate of USD/CNY with interaction of crucial indicators by using the Long short-Term Memory (LSTM) model.

1.4 Scope and Limitation

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The FOREX financial market can be confirmed as the most variable and biggest market around the world.Although there are many methods and models to analyse and forecast the FOREX market. Because of the unpredictability and sensitivity of its characteristics for macroeconomic factors , it is significantly challenging to apply accurate forecasting of the futural values . However, with its intensive time-series relevance ,recently, there are developing loads of advanced techniques of ML (Machine learning) ,which erect more steadfast and precise predictive models such as LSTM(Long Short-Term Memory) and ARIMA(Autoregressive Integrated Moving Average) ,to address the accuracy issue of the FOREX predicting . Through the overall literature,it demonstrates that identifying the key features about the FOREX such as CPI(consumer price indices),interest rate and 10-year bond rate. Additionally, the project assimilates insights from EDA(exploratory data analysis) , feature engineering and results from modelling.

2.2 Macroeconomic Indicators

Macroeconomic indicators play a crucial role in foreign exchange rate movement,such as CPI,interest and 10-Year bond.There are some cases to interpret the relationship among these factors.For an instance,Kim and his mates (Kim,et al,2022) fathered the inadequacy of traditional econometric models, such as the uncovered interest parity (UIP) and purchasing power parity (PPP), which exceed the random basic models. These indicators can be confirmed as the key dataset for machine learning processing to forecast the future foreign exchange rate.

2.2.1 Interest Rate

Some studies have interpreted how interest rate movements impact on exchange rate over the time.For an instance,as the display of Yuenan(2023)'s study, while a central bank of the country increases their interest such as the Federal Reserve in the United States, it will lead to the result that rising their currency ,because international investments will benefit from high interest level,and then in the financial market, the demand of purchasing USD regularly rise.This study also declaim that because of interest highly boosting, the Chinese Yuan has devalued significantly.Through his study which discussed and implemented the impact of Federal Reserve System raising interest rate on the exchange rate of the US Dollar against the Chinese Yuan ,with the analysis of ARIMA,it illustrated that hiking the interest rate of the Federal Reserve led to the exchange rate of pairs USD/CNY steadfast increased over the time.

In addition,according to Marcin Kolasa's study(Marcin Kolasa,et,2022), they added behavioral agents to stretch the standard open economy New Keynesian framework .It conveyed a theory that the exposing to interest rate parity condition in a way and consisting with recent testing-evidence reconsidered that conducting positive monetary policy such as boosting the interest by central bank can significantly influence international output

comovement positively.Moverover,they developed the New Keynesian model with appraising an extension to successfully mitigated many problems which demonstrated the relationship between exchange rates and interest rates.Through their improvement of the model,they observed that the trending of the exchange rate under two different monetary policy statuses:Conventional Monetary Shocks and Low for Longer Policies.In the final exhibition,the result showed that during these statuses,the trending of exchange rate was intensively affected by the interest movement.

2.2.2 CPI

CPI usually can display a consumer situation of a country or place,and also is a key economic indicator of consumer price inflation. The CPI provides effective insight into how inflationary pressures may impact market patterns, company financial tables, and stakeholders. Joseph and his mate(Joseph,et 2023) researched the relationship between exchange rate and consumer price index with data context from South African.They adopt OLS (ordinary least squares) regression with the aid of Gretl econometric and statistics software to analyse the South Africa 2022 Consumer Price Index (CPI) YoY.After modeling ,they found that two-month later, there are a possible and positive movement in exchange rate while consumer price index get alteration. Furthermore, offering policies of economy and finance that may cushion the exchange rate volatility could catalyse a favourable effect on domestic prices, which is direly needed for the economic good of the citizens. They summarized their project and concluded that the oscillation of South African currency may positively and significantly impact the consumer price index,which means that there is a powerful relevance between CPI and FOREX.

There is another study to demonstrate the powerful relationship between CPI and foreign exchange rate. Through the study of Didarul and his mates (Didarul.et 2024),it is not herculean to recognise the dynamics of Uruguay's real effective exchange rate with the aid of the MundellFleming model. They analyzed crucial economic indicators and employed econometric techniques, and therefore revealed the significant insights of the determinants of Uruguay's currency. With

the analysis of the short-term real effective exchange rate in Uruguay currency,they deployed the Ordinary Least Squares(OLS) regression model.The model formula is below:

$$REER=\beta 0+\beta 1\times USLR+\beta 2\times M2+\beta 3\times CPI+\beta 4\times WIR+\varrho \quad (2.1)$$

After processing,they concluded that with maintained appropriate CPI,it will promote Uruguay currency healthily.

2.2.3 10-Year Bond

10-year government bonds play a significant role in affecting FOREX markets,Oscillations of these bonds have a great impact on currency valuations through various factors such as interest rate differentials,capital flows and stakeholders risk.Recently,there are some academic studies to elucidate these facts.Usually,10-Year bond indirectly impact FOREX market by directly influenced interest rate.Although it can not directly affect FOREX market,it still can be confirmed as a significant factor in short time.

As the study of Rosnawintang and his mates(2019),they analyzed the relationship between Indonesian currency and bond yield.According to their study ,it is a crucial theory that domestic interest rate is the sum of international interest rate and expectation of depreciation (or appreciation) of domestic currency exchange rate against foreign currency .Through their VAR(Vector autoregression) model,they collected the data of IDR/USD exchange rate and bond yield from January 2006 to December 2018 with monthly time series.After the feature engineering,they concluded that it is effective that the causal relationship between the government bond yield and paires IDR and USD exchange rate during the short-term relationship.They also declaimed that in the first three months, the response of the government bond yield against the exchange rate between IDR and USD was very strong (significant 1%).

2.3 FOREX prediction model

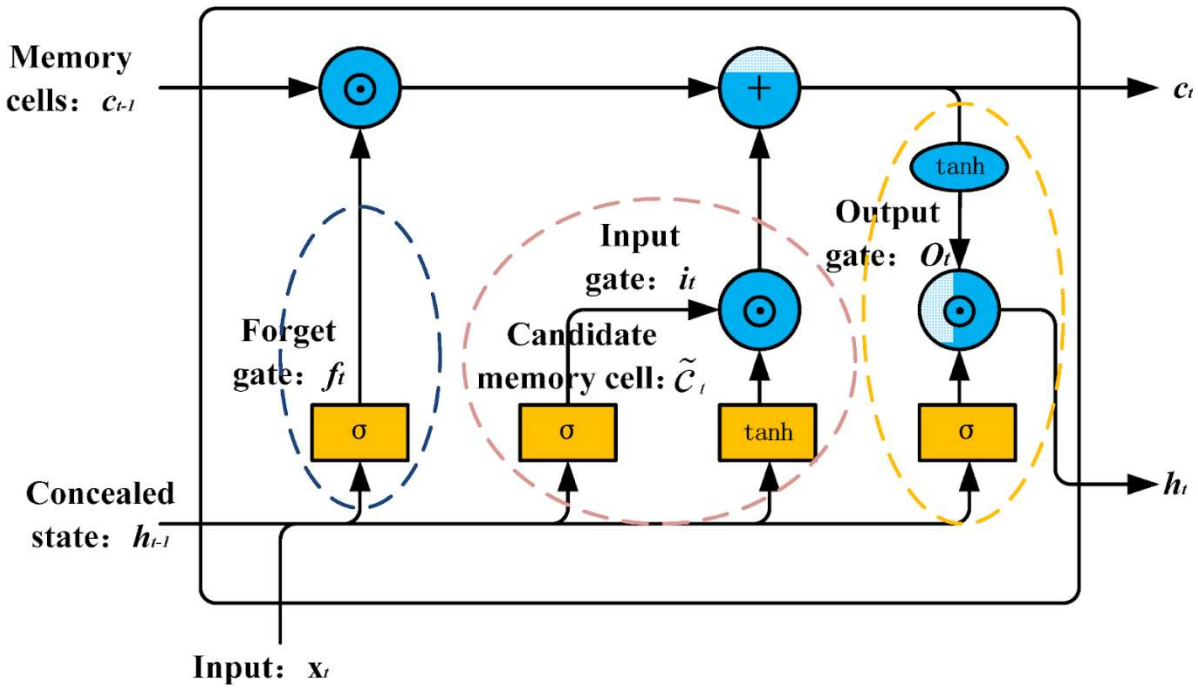
There is a famous deep learning technique which is called “long short-term

memory" (LSTM).It is known as very efficient in many time-series forecasting problems, to navigate direction predictions in the FOREX market. Going through recent studies,for predicting the future exchange rate,it is certain that many researchers prefer the LSTM model to complete the forecasting work.A study by (Deniz et al.,2021)stated that the single LSTM models were vaguely better than ME_TI_LSTM in profit_accuracy, which was less than 1% during the one-day size.However,in three-day scale, the individual LSTM models outperformed the ME_TI_LSTM in profit_accuracy by 5.81%.In the five days ahead forecation, they found that in the profit_accuracy results,it illustrated the similar relevance of individual LSTMs and the ME_TI_LSTM .

2.3.1 Long Short Term Memory Model (LSTM)

Long Short-Term Memory (LSTM) networks, a type of Recurrent Neural Network(RNN), were designed to resolve issues present in traditional RNNs, such as the vanishing and exploding gradient problems (Y. Bengio,1994). LSTM has three key characteristics :memory cells,gates mechanism and cell state.In terms of memory cells ,it is for storing and managing information around the time series,ensuring maintenance of relevant data while dropping irrelevant data.As for gates mechanism,usually,there are three different gates in the LTSM for regulating the flow of information.First one is called Forget Gate,which determining what data from cell state will be dropped .The second one is named Input Gate,storing the effective data in the cell state.The third one is called Output Gate ,picking the valued output to export at each step.The cells state can be imagined by a pathway,which engaged through the whole sequence, supporting data to flow with minimal modification unless regulated by the gates.The flow of LSTM details shows below Figure 2.1.

Figure 2.1 the work flow of LSTM



Based on the framework of LSTM, nowaday,researchers implement a series of applications. For example,because of an intensive allergy to time series,it is effective to predict the trending of the stock prices, weather, future FOREX rates and so on.It can not be ignored that LSTM plays an important role in processing data with its advantage.Speaking about advantages,contrasting to the traditional RNNs,LSTMs can deal with learning dependencies over long sequences well.Moreover,it also can accept the various length of sequences and convey a suitable data to tasks with irregular time series.

It performs well in alleviating the vanishing gradient problem,ensuring training operation stably around the processing.

2.3.2 Case Studies

Because of the importance of the FOREX,it is obvious to see more and more researchers involving detecting the future trending of exchange rate.By the paper (Ling Qi et al.,2020),they applied LSTM model in predicting the price of exchange rate with using technical indicators as the feature.Firstly,they collected

the data including four major currency pairings : GBP/USD, EUR/GBP, AUD/USD and CAD/CHF and 15 minutes interval data through Jan 2005 to Sep 2017 for training and testing data from Oct 2017 to Sep 2020 .Secondly,they set a sets of standard regression metrics which are mean square error, root mean squared error,mean absolute error and mean absolute percentage error.Then they select cleaned data for entering the LSTM model .At the final phrase of the experiment,they found that the performing of LSTM and GRU models substantially exceed th RNN model for three currency pairs – EUR/GBP, AUD/USD and CAD/CHF.In addition,they identified that concerning different time series,currency pairings perform the difference simultaneously.

Another study also highlighted the effectiveness of the LSTM in applying FOREX prediction.According to the research (Hasnaet al.,2020),they gathered EUR/USD historical data from 4 May 2010 to 19 October 2020 with 2718 records of daily time frame data and 64897 records of 1-hour timeframe data.For getting the best hyperparameters to process in the LSTM model ,they stored the RMSE test data from 02 January to 19 October 2020. After examining the daily timeframe,1-hour timeframe and comparing the previous years basing on 2020 prediction ,they found that the best hyperparameters were 2 hidden layers and 10 neurons with a dropout layer (0,1 rate) and the RMSE result which meant 0,624 × 10 in in the daily timeframe.In addition,as for 1-hour timeframe,it was 1 hidden layer and 5 neurons without a dropout layer and the RMSE result which meant 0,135 × 10 .They concluded that LSTM model process more accurate and effective than with chaos and turbulence such as 2020 COVID-19.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

The FOREX(foreign exchange) market faces individual challenges for predicting because of its unique characteristics such as dynamic,non-linear,and fiendishly unstable.Therefore,using traditional approaches often weakly confront capturing these problems,applying advanced computational techniques.This project implements tackling these challenges and promoting the accuracy of FOREX predictions with the use of a machine learning model,LSTM(Long Short-Term Memory) .

In this project,adopting a structured methodology to transform raw financial data into valuable and actionable insights. First of all,collecting and seeking comprehensive and high-quality datasets which include historical exchange rates and additional indicators of macroeconomic and microeconomic.Then ensuring analyzing clean,consistent and veracity data,it is a crucial procedure to beforehand process data for preparation.Conducting engineering techniques to effectively extract meaningful and valuable patterns and relationships for facilitating the predictive power of the model.

The following subsections demonstrate every step of the methodology ,which is from data source description,data Collection and analysis to model evaluation.With the use of this structured approach,the project transforms raw ,non-pattern and complex data from the FOREX market to clear,structured and

valuable data for stakeholders,and offers a set of forecasting insights.

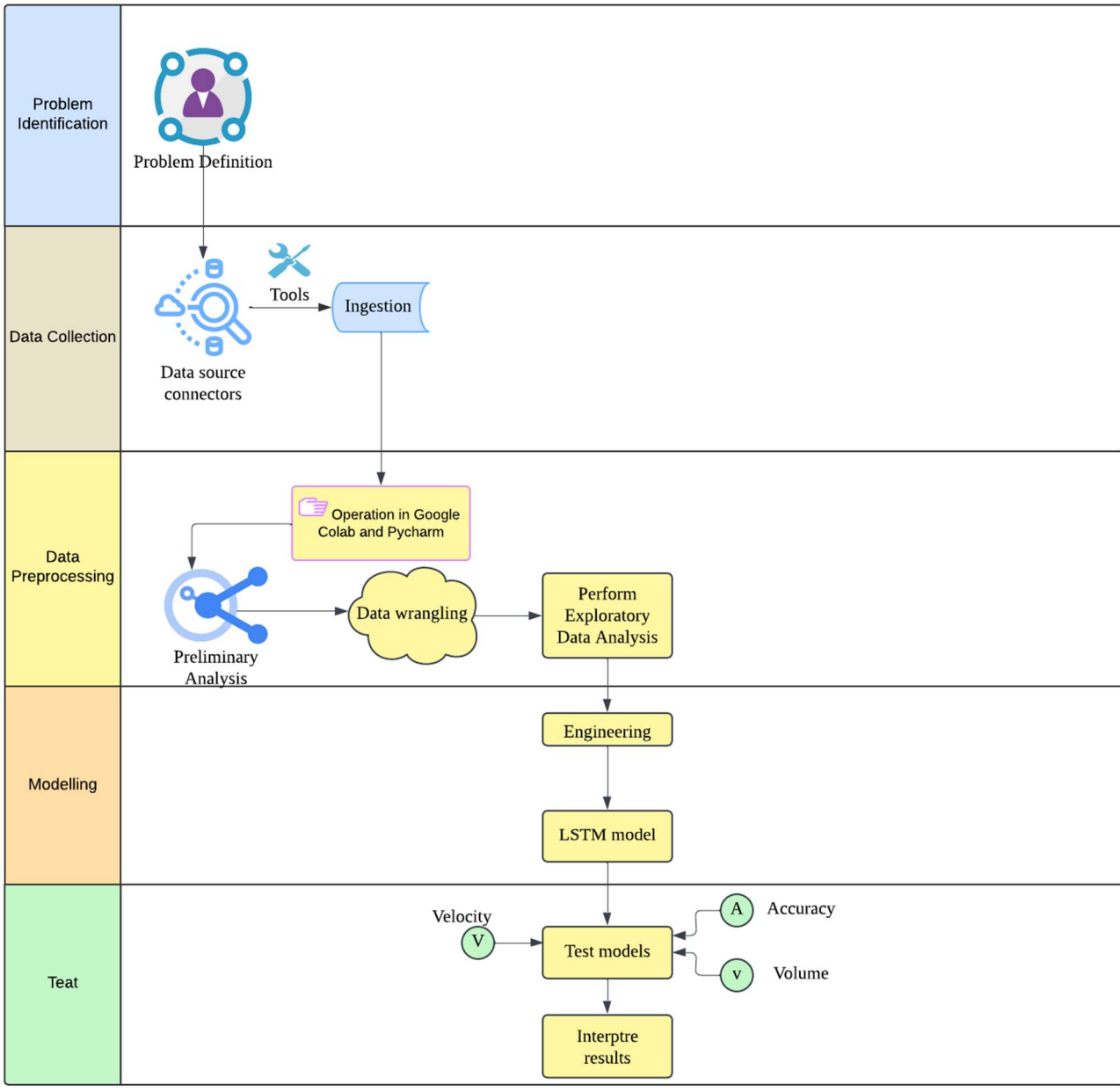


Table 3.1 Project Research Framework of FOREX prediction

3.2 Data Collection

In this project,it is used as a dataset from the website of Exchange Rates UK for pairs USD/CNY.In the meantime,also it is confirmed that dataset from the

website of Alpha Vantage for pairs USD/CNY. The range of both Dataset are from 28th December 2014 to 27th December 2024.The size of the first one is 168 KB,and the second one is 112 KB.Due to improving the accuracy of the forecasting about the pairs USD/CNY,there are three additional factors which are 10-year Treasury yield,interest and inflation from US and China to need to concern.So,the dataset of 10-year Treasury yield derived from the Alpha Vantage ,and Chinese one is from People's Bank of China (PBOC).As for CPI data of US and China,they are both from IMF..Moreover,the dataset of US interest rate is from the Alpha Vantage while China one is from PBOC.It shows below Table 3.2

Datasets	Attributes
China_10YBond.xlsx	Date: the time of the yield observation Rate: the bond interest rate
Exchange_PairsUSD&CNY_Fromexchangerates.xlsx	Date:the time of the exchange rate observation US Dollar to Chinese Yuan:the value of 1 USD in Chinese Yuan (CNY)
CPI_China.xlsx	Date:the time of the CPI Consumer Price Index:the raw CPI value for the given time era
CPI_US.xlsx	Date:the time of the CPI Consumer Price Index:the raw CPI value for the given time era
China_1YLPR.xlsx	Date:the time of the LPR observation 1 Y: interest rate for short-term loans

Table 3.2 Four factors data(10-year Treasury,interest and inflation

3.3 Data Pre-processing

Due to massive data ,it is quite crucial to beforehand process these various datasets,ensuring later machine learning model analysis works smoothly.It transform raw,non-pattern and chaotic data to clear,structured and comprehensible data for machine learning modeling.Below diagram illustrates the entire steps of Data Pre-processing.

3.3.1 Preliminary Analysis

Preliminary analysis plays an important role in data pipeline .Through the eight datasets which are CPI_US, CPI_China, BOP_China, BOP_US,interest_US ,interest_China,10YTreasury_China,10YTreasury_from China and US,for processing in mode stably,it is necessary to merge these eight datasets into one.Checking the information of these datasets by Figure 3.1 ,it is essential to unify the format of the date,which transforms Quart,daily to monthly.

Figure 3.1 realtime exchange rate between China and US

	timestamp	open	high	low	close
0	2024-12-30	7.2974	7.2997	7.2967	7.2994
1	2024-12-26	7.2973	7.2988	7.2955	7.2973
2	2024-12-25	7.2972	7.2983	7.2968	7.2972
3	2024-12-24	7.2946	7.2984	7.2946	7.2946
4	2024-12-23	7.2970	7.2981	7.2943	7.2970
...	...	...	...	...	...
2601	2015-01-04	6.1961	6.2090	6.1961	6.1961
2602	2015-01-01	6.1961	6.1961	6.1961	6.1961
2603	2014-12-31	6.1961	6.1961	6.1961	6.1961
2604	2014-12-30	6.1920	6.1964	6.1810	6.1920
2605	2014-12-29	6.2125	6.2218	6.1863	6.2125

[2606 rows x 5 columns]

3.3.2 Data Cleaning

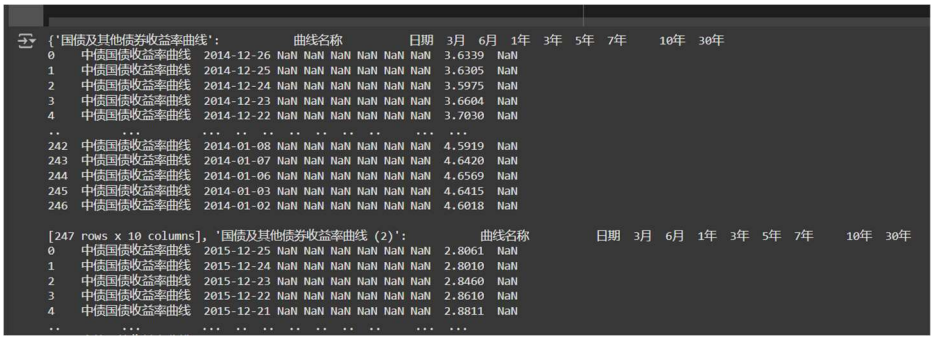
Data Cleaning is a significant step of the Data science analysis that engages a set of preparation of raw data for later machine learning modeling.It significantly implements the goals which is to attain a veracity ,complete,and consistent data.In addition, it would facilitate the quality of insights and predictions.From figure 3.2,it is clearly to showcase the circle flow of the Data Cleaning.

Figures 3.2 Data cleaning cycle

First of all,according to the above circle of Data Cleaning steps,dealing with the 10-year Treasury yield of the US and 10-year Treasury yield of China.Hading missing data of these two datasets is the first step.Checking Figure 3.3 ,it is straightforward to find the messorder of this dataset,China_10YBond.xlsx, missing value ,extra columns and language format.So,using pandas tool to tackle these problem for ensuring appropriate data order .After cleaning,Figure 3.4

displays that the table is transformed to the valid data table.

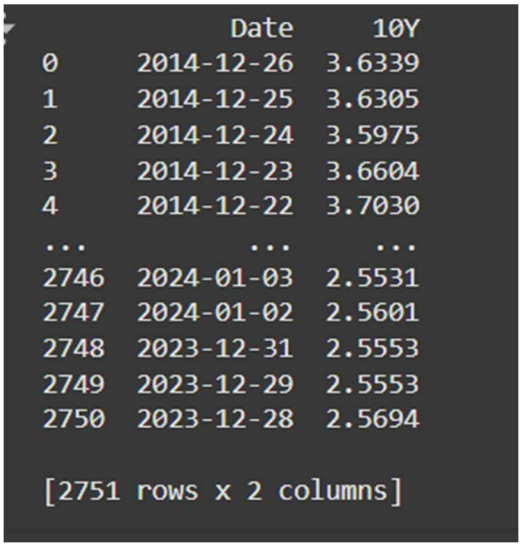
Figure 3.3 Raw Chinese interest data



```
{'国债及其他债券收益率曲线':          曲线名称         日期  3月  6月  1年  3年  5年  7年    10年  30年
0    中债国债收益率曲线  2014-12-26 NaN NaN NaN NaN NaN NaN  3.6339  NaN
1    中债国债收益率曲线  2014-12-25 NaN NaN NaN NaN NaN NaN  3.6305  NaN
2    中债国债收益率曲线  2014-12-24 NaN NaN NaN NaN NaN NaN  3.5975  NaN
3    中债国债收益率曲线  2014-12-23 NaN NaN NaN NaN NaN NaN  3.6604  NaN
4    中债国债收益率曲线  2014-12-22 NaN NaN NaN NaN NaN NaN  3.7030  NaN
..          ...  ...  ..  ..  ..  ..  ..  ..  ..     ...  ...
242  中债国债收益率曲线  2014-01-08 NaN NaN NaN NaN NaN NaN  4.5919  NaN
243  中债国债收益率曲线  2014-01-07 NaN NaN NaN NaN NaN NaN  4.6420  NaN
244  中债国债收益率曲线  2014-01-06 NaN NaN NaN NaN NaN NaN  4.6569  NaN
245  中债国债收益率曲线  2014-01-03 NaN NaN NaN NaN NaN NaN  4.6415  NaN
246  中债国债收益率曲线  2014-01-02 NaN NaN NaN NaN NaN NaN  4.6018  NaN

[247 rows x 10 columns], '国债及其他债券收益率曲线 (2)':          曲线名称         日期  3月  6月  1年  3年  5年  7年    10年  30年
0    中债国债收益率曲线  2015-12-25 NaN NaN NaN NaN NaN NaN  2.8061  NaN
1    中债国债收益率曲线  2015-12-24 NaN NaN NaN NaN NaN NaN  2.8010  NaN
2    中债国债收益率曲线  2015-12-23 NaN NaN NaN NaN NaN NaN  2.8460  NaN
3    中债国债收益率曲线  2015-12-22 NaN NaN NaN NaN NaN NaN  2.8610  NaN
4    中债国债收益率曲线  2015-12-21 NaN NaN NaN NaN NaN NaN  2.8811  NaN
..          ...  ...  ..  ..  ..  ..  ..  ..  ..     ...  ...
```

Figure 3.4 cleaned Chinese interest data



	Date	10Y
0	2014-12-26	3.6339
1	2014-12-25	3.6305
2	2014-12-24	3.5975
3	2014-12-23	3.6604
4	2014-12-22	3.7030
...	...	...
2746	2024-01-03	2.5531
2747	2024-01-02	2.5601
2748	2023-12-31	2.5553
2749	2023-12-29	2.5553
2750	2023-12-28	2.5694

[2751 rows x 2 columns]

3.3.3 Data Merging

For smoothly and beforehand processing in the analysis and later modeling,it is quite necessary to integrate all key data frames into one comprehensive dataframe.So,using pd.merge() of pandas package is to combine df_FOREX,df_10Y_US, df_cpi_us,df_cpi_china,df_interest_US,df_interestChina and df_10Y_China,and then it will get a overall dataset:df_whole_frame,which is illustrated in Figure 3.5 below.

Figure 3.5 Merging among data frames

```
import numpy as np
#create a dataframe with date from 2014-01-01 to 2024-12-31
import pandas as pd
start_date = '2014-01-01'
end_date = '2024-12-31'
date_range = pd.date_range(start=start_date, end=end_date, freq='D')
df_date = pd.DataFrame({'Date': date_range})

#print(df_date)
#merge df_FOREX and df_date
df_date['Date'] = pd.to_datetime(df_date['Date'])
df_FOREX['Date'] = pd.to_datetime(df_FOREX['Date'])
df_interestChina['Date'] = pd.to_datetime(df_interestChina['Date'])

# Now perform the merge
df_whole_frame = pd.merge(df_date, df_FOREX, on='Date', how='left')
#fill NaN by using forward
df_whole_frame['USD_rate_CNY'] = df_whole_frame['USD_rate_CNY'].fillna(method='ffill')
```

	Date	USD_rate_CNY	10YTreasury_US	CPI_US	CPI_China	\
0	2014-01-01	6.0540	3.00	107.273607	113.313217	
1	2014-01-02	6.0507	3.00	107.273607	113.313217	
2	2014-01-03	6.0515	3.01	107.273607	113.313217	
3	2014-01-04	6.0515	3.01	107.273607	113.313217	
4	2014-01-05	6.0515	3.01	107.273607	113.313217	
...	...	...	...	...	...	
4017	2024-12-27	7.2950	4.62	144.684725	132.036865	
4018	2024-12-28	7.2950	4.62	144.684725	132.036865	
4019	2024-12-29	7.2950	4.62	144.684725	132.036865	
4020	2024-12-30	7.2994	4.62	144.684725	132.036865	
4021	2024-12-31	7.2994	4.62	144.684725	132.036865	

	interest_US	interest_China	10Y_China
0	0.07	5.73	4.60
1	0.07	5.73	4.60
2	0.07	5.73	4.64
3	0.07	5.73	4.64
4	0.07	5.73	4.64
...	...	...	...
4017	4.48	3.10	1.69
4018	4.48	3.10	1.69
4019	4.48	3.10	1.69
4020	4.48	3.10	1.69
4021	4.48	3.10	1.69

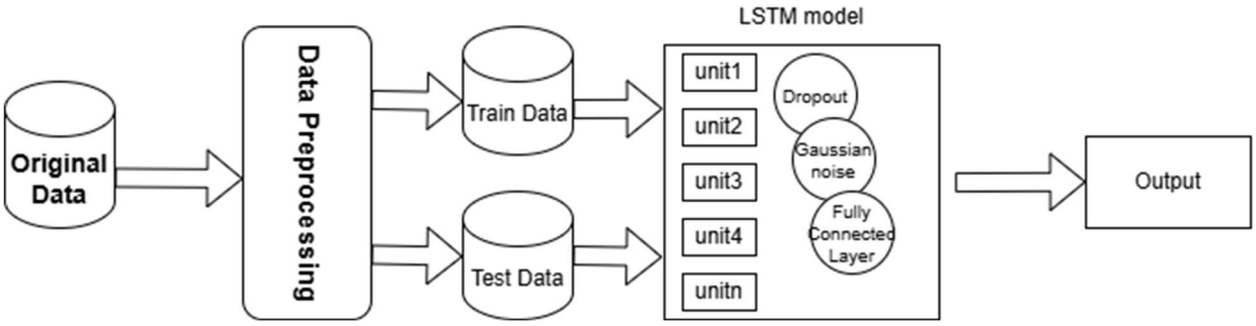
3.4 Data Modelling

In this project, using the LSTM model as a major analysis tool. Long Short-Term Memory (LSTM) networks is a special form of the RNN (Recurrent Neural Network) architecture,which is for facilitating the performance of a neural network with past data inputting and solving some traditional RNNS problems such as vanishing gradient.From the Figure 3.4.1,it is understandable to figure out the flow of LSTM working when integrating LSTM into Data science life circle.Firstly,preparing dataset with the characteristic of time series correspond the structure of LSTM.During the Data Preprocessing,it is significant to normalise the data under the standard of LSTM.When the Data enter into model circle,LSTM will work stably and appropriately with these processed data.

Through a set of crucial steps of the Data Preprocessing,splitting the processed data to two same datasets for the accurate forecation of LSTM.One named Train Data and another named Test Data .Within the processing of LSTM model,it will

implement three important phrases:Dropout ,Gaussian noise and Fully Connected Layer.During th training ,Dropout plays a role in preventing overfitting by memorising long-term dependencies.After the dropout phrase,the data will go through the Gaussian Noise ,which ensures that the inputs makes the LSTM model more steadfast to noisy.After passing the Gaussian Noise phase,Fully Connected Layer will be the final one of the circle.And it will summarize the outputs from the LSTM to offer the expected values.

Figure 3.6 LSTM working flew



CHAPTER 4

INITIAL RESULTS

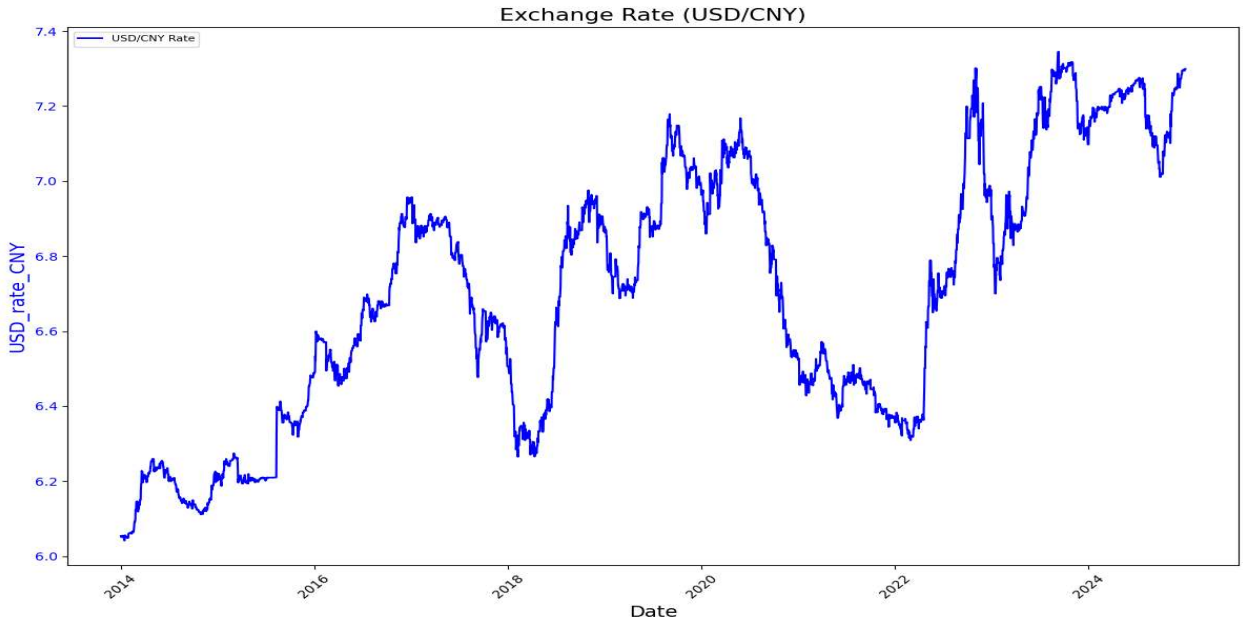
4.1 Data Visualizations

4.1.1 Interest Indicator

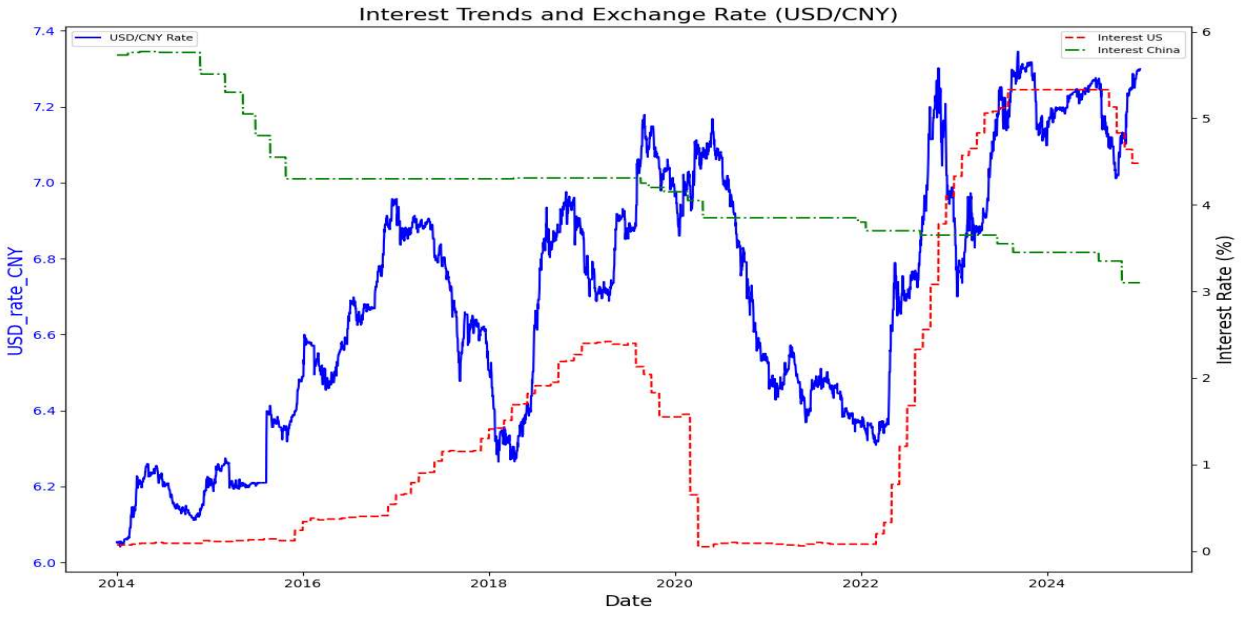
It is meaningful to overview the trending of USD/CNY through the line chart,which is Figures 4.1.It is straightward to see that the trending of line almost incline upwards.However,there are two bottomest era which are around 2018 and around 2022.Then,it is necessary to concern the major reasons which lead to this result.So ,involving with the trending of the difference between US interest and China interest,which are majorest factors to affect the trending of the FOREX ,USD/CNY paires.

When considering the interest influence,it will be able to demonstrate the the trending of pairs USD/CNY.From the Figure 4.2 to see,it is so clear to figure out the explanation from 2020 to 2022 .Because the Chinese interest was much higher than the US's which led to CNY gets strong increasing contrasting with USD , and the market expected Federal Reserve would decrease the future interest rate,the trending of USD/CNY intends to show declining over the years.

On the other hand,from 2023 to 2024,due to the continually boosting interest rate by the Federal Reserve tightening policy,the trend started to turn upward from the bottomest point.It is straightward to note the interest rate of US was higher than the Chinese one by almost 2% in a short time.Through the whole period,it illustrated interest played a significant role in the movement of the FOREX directly.



Figures 4.1the Trending of USD/CNY



Figures 4.2 the interest trending impact the exchange rate

4.1.2 CPI Indicator

CPI usually can genuinely reflect the goods price trending of a country,and is a concerning factor whether the central bank or policy maker decides to increase the interest or not.Around the global central bank ,specially for advanced economic countries,they have to control CPI fluctuation between 2% to 5%.It will urge policy makers to employ some adjustments when the CPI works out of range.

Figure 4.3 illustrates the fluctuation of CPI and exchange rate through the period of 2014 until 2024.Generally, both curves are intending to turn upward in long time because central banks need to maintain the 2% CPI goal to ensure domestic economy runs healthily.From 2020 and 2024,the CPI of US presents consistent upward movement,suggesting an increasing in value.On the other hand,Chinese one presents quite stable and slow increasing.From 2020 to 2024,it illustrates almost flat curve ,comparing with US's one.Observing the details of Chinese CPI,it may make policy maker intending to lower the interest for avoiding the risk of deflation.

To better demonstrate the relationship between CPI and FOREX ,involving interest rates which reflect the actual adjustment of central banks is quite effective .As Figure 4.4 which showcases the fluctuations in CPI ,interest rates and exchange rate between China and US,it is clear to see what happened from 2020 to 2024.There is a intensive drop of US interest around 2020 because the Federal Reserve implemented a quiet peaceful policy for due to COVID19 damage the system of economic operation.However,quick decreasing interest in short time and the global health event brow up the CPI together.Definitely,it led to the Federal Reserve employed a series of tightening policy to lower the temperature of CPI.Therefore,the result shown a sharp growth of the US interest in 2023 , and meantime the paires USD/CNY was experiencing abrupt rise.

So,CPI can not directly force the movement of the FOREX,but CPI can significantly affect the changes of interest rate.Then interest rate directly impacts

on the FOREX market.

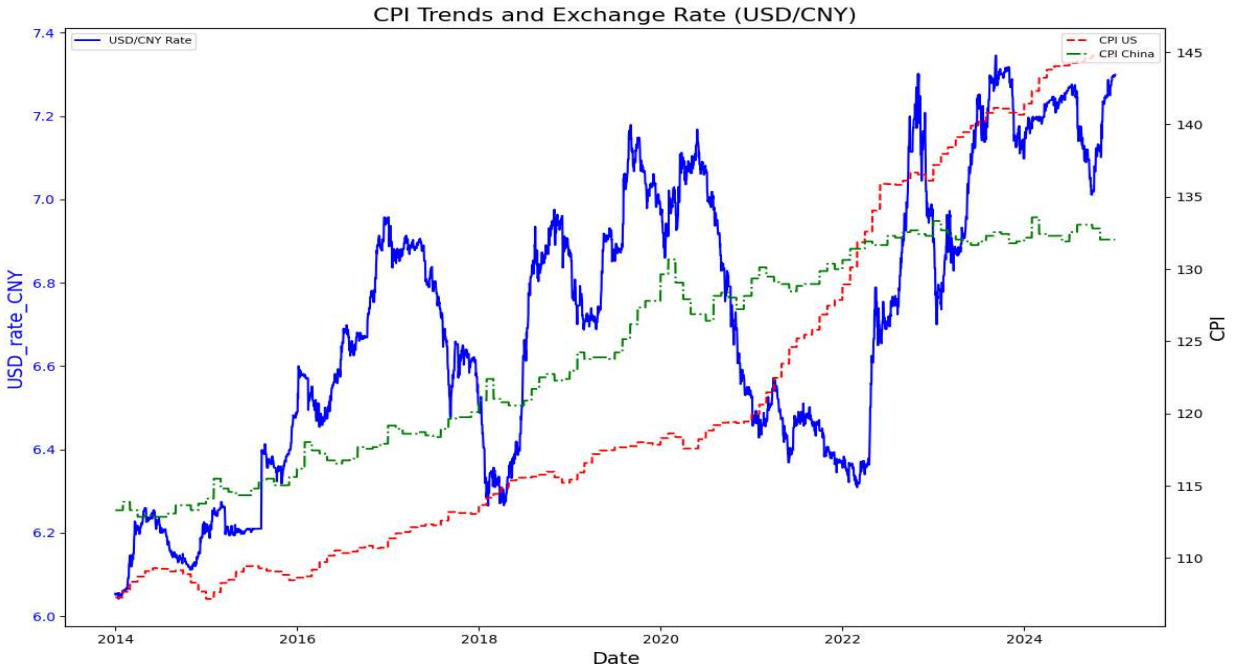


Figure 4.3 CPI and exchange rate trending

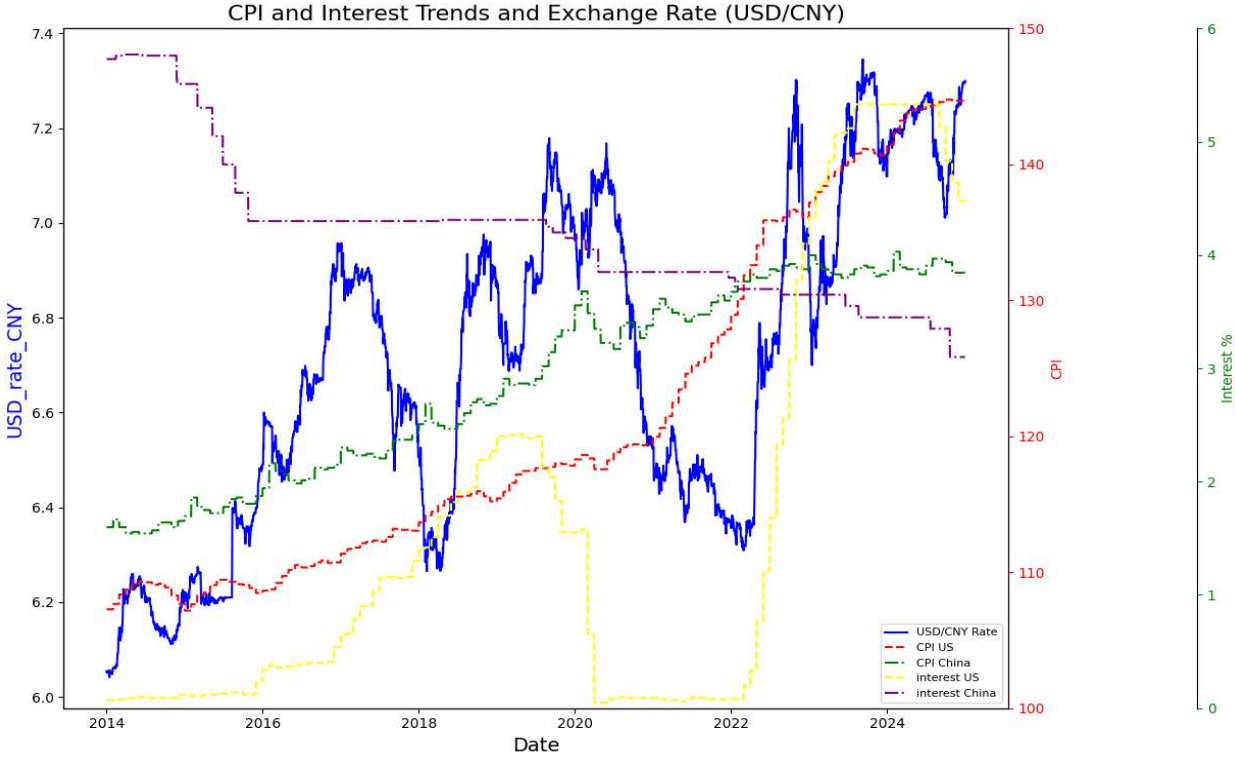


Figure 4.4 CPI, Interest and exchange rate trending

4.1.3 10 Years Bond revenue Indicator

In the financial market,government debts usually were confirmed as non-risk financial assets.Specially,10-year bond yield often is verified as the standard of concerning various interest rates such as short-term and long-term loan market .It is a effective metric to reflect the expectations of investor and stakeholders about future inflation, economic growth, and interest policies.There is a interesting market feedback mechanism,which means that it will imply higher inflation or stronger economic growth is coming when revenue of the 10-year bond is rising.

Through Figure 4.5,it presents the trending of the exchange rate under the influence of a 10-Year bond around the 10 years.From 2014 to 2024,the curve of exchange rate shares similar direction with the curve of US ten years national bond.At the China side,over the observed period,it generally illustrates the steadily decline.Specially,from 2021 to 2024,there is a noticeable and consistent downward trend.It can assume that market expects Chinese central bank will possible continually execute decline interest rate policy in the future.

When analysing the relationship between interest and 10-year bond ,the Figure 4.6 showcases that the trend of US interest rate and the trend of the US 10-year Treasury have almost the same pattern around the 10 years.From 2020 to 2024,the fluctuation of 10-year bond always leads the trend of the interest rate.It also presents the Federal Reserve often applies dramatical interest policy around the period through the quick upward and downward oscillations,contrasting to the flat and steady of Chinese one.

On the Chinese side,10-year bond steadily decline as the result that the consistent lower interest rate policy around the period.The 10 years national debts market is reflecting the trend of the Chinese central bank policies in the future.Merging the interest,10-year bond and exchange rate,it states that they play a significant role in movement of exchange rate.

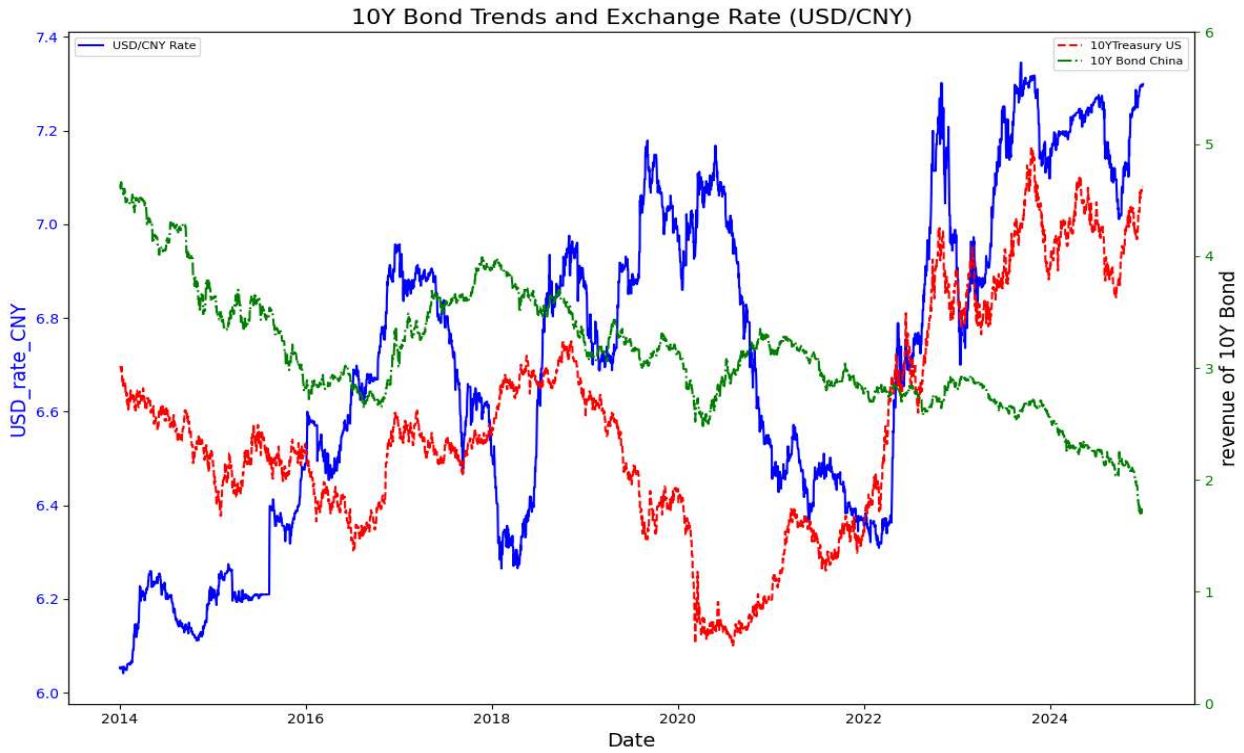


Figure 4.5 10Y Bond and USD/CNY trending

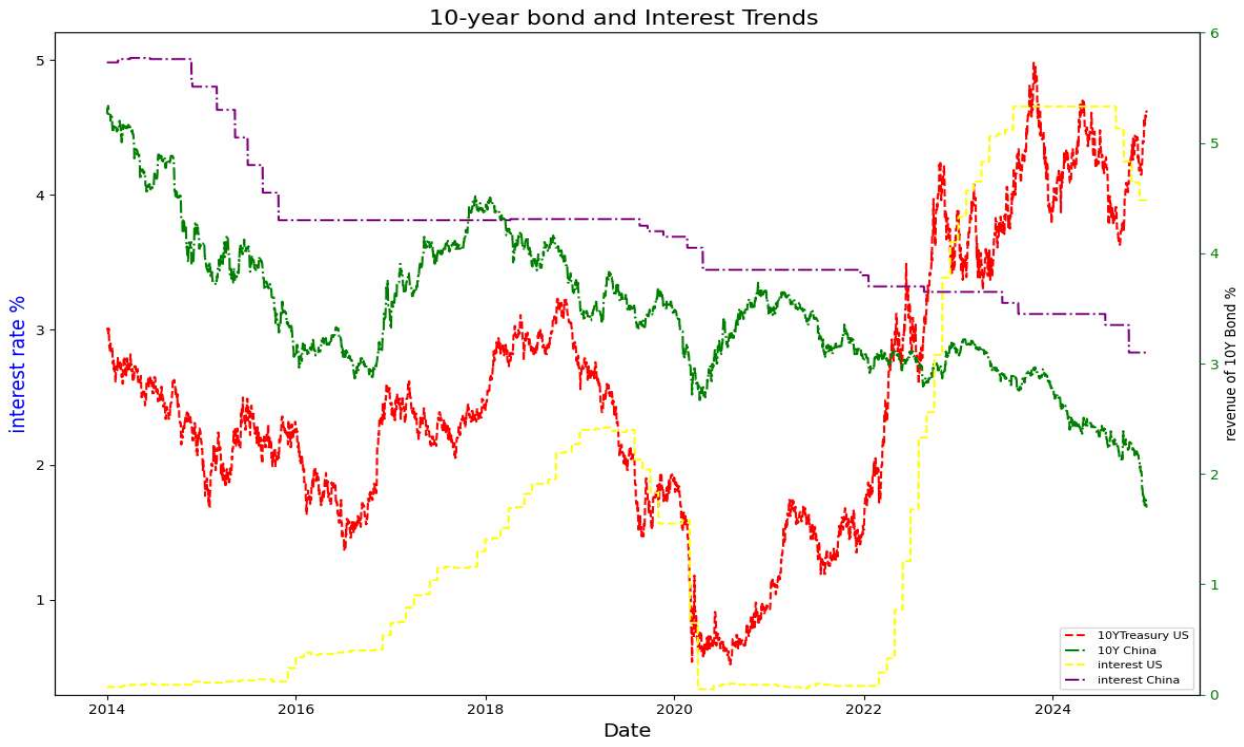


Figure 4.6 10-Year bond and interest rate

4.2 Descriptive Analysis

The below table offers a complete description of several financial factors over the 10 years,including exchange rate between US and China, CPI_China, CPI_US, Interest_US, Interest_China ,10Y_China and 10Y_US.There are 8 different statistical metrics that were presented.The 8 metrics are displayed for each distinct financial indication which are count,mean,minimum(min),25th percentile(25%),median(50%),75th percentile(75%) ,maximum(max) and standard deviation(std).

The dataset of this project consists of 4022 observations for individual financial variable,suggesting a stalwart sample size which fits for meaningful statistical analysis.The average exchange rate between China and the US is approximately 6.755, while the median is 6.7724, impling a slight skew which directs the lower end of the range.Also, the values fluctuate between 6.1775 and 7.3278, while the standard deviation is just 0.3132, offering the fairly steadiness of the exchange rate over the 10 years.

By the US CPI, it has an average of 121.17 and a median of 117.52, and it presents that value oscillates between 107.18 and 144.76.Moreover, the standard deviation shows value 12.18 confirm a middling variability, providing a significant inflationary trend.On the contrary, the average CPI of China is quite a little higher than US ,which it is 124.11.And its median is 124.23 with a narrow range from 112.85 to 133.58.However ,Chinese standard deviation is only 6.94, indicating more standfast , compared to the US CPI.

Speaking about Interest Rate, the US average one is 1.63, and its median arrives 0.85, displaying that most observations are at the lower the average. The range distributes from 0.05 to 5.33, but the standard deviation shows high value reaching 1.84, resulting in unstable fluctuations, probably because of energetic monetary policies. In contrast, the Chinese one exhibits intensive resilience and is

more stable, with a mean of 4.2 and a median of 4.3. The values range from 3.1 to 5.77, and the standard deviation of 0.65 suggests minimal movement.

As for the 10-Year Bond Yield for China,its average value is 3.13, with a median of 3.09 and moves between 1.69 and 4.66. The standard deviation is 0.54, displaying limited fluctuation. Meanwhile, the 10-Year Bond Yield for the US is lower on average at 2.49, with a median of 2.36. The values change from 0.52 to 4.98, yet standard deviation is 1.00 which is higher than Chinese, providing more volatility in the US bond market.

Overall, the project data reveals crucial differences between the US and Chinese economic indicators. The US exhibits more variability in interest rates, bond yields, and inflation, presenting a dynamic and potentially active monetary environment. In contrast, China's indicators reflect more stability due to narrower ranges and lower variability in interest rates and bond yields. These patterns probably suggest the different structures and policies of economics between the two countries.

Table 4.1 Descriptive Analysis of Factors

statistic	Exchange rate	CPI_US	CPI_ China	Interest_ US	Interest_ China	10Y_ China	10Y_US
count	4022	4022	4022	4022	4022	4022	4022
mean	6.755219	121.170703	124.109282	1.628834	4.200144	3.132277	2.486363
min	6.177500	107.177760	112.853524	0.050000	3.100000	1.690000	0.520000
25%	6.473700	110.684207	117.680253	0.100000	3.700000	2.780000	1.780000
50%	6.772400	117.523291	124.230807	0.845000	4.300000	3.090000	2.360000
75%	7.016100	132.401040	131.588196	2.390000	4.310000	3.500000	2.980000
max	7.327800	144.763145	133.575159	5.330000	5.770000	4.660000	4.980000
std	0.313176	12.179279	6.942263	1.844476	0.654749	0.538024	1.004274

4.3 Initial Insights

To predict the future value of pairs USD/CNY by using an LSTM model, the LSTM analyzes the historical patterns of five key factors: CPI_US, CPI_China, Interest_US, Interest_China, and 10-year bond yields. Then integrating these variables seamlessly into a deep learning pipeline,the model transforms raw data into an effective predictive value. The below Figure 4.7 showcases trend of the USD to CNY exchange rate over the next 30 days, originating from historical data and additional features.

From the graph, it can probably be concluded that over the next 30 days, the USD displays a strong and dynamic influence in the FOREX market against the Chinese Yuan. By contrary, the Chinese currency interprets a consistent weakening trend in the financial market,because of a potential decreasing interest rate environment.Therefore,this result will may lead to Chinese policy maker hedges the risk of exchange rate by applying a series of financial tools ,ensuring the steadiness of the CNY .The upward curve provides a dominant USD position,emphasizing its resilience and muscle in the global forex market due to the Federal Reserve tightening interest policy.

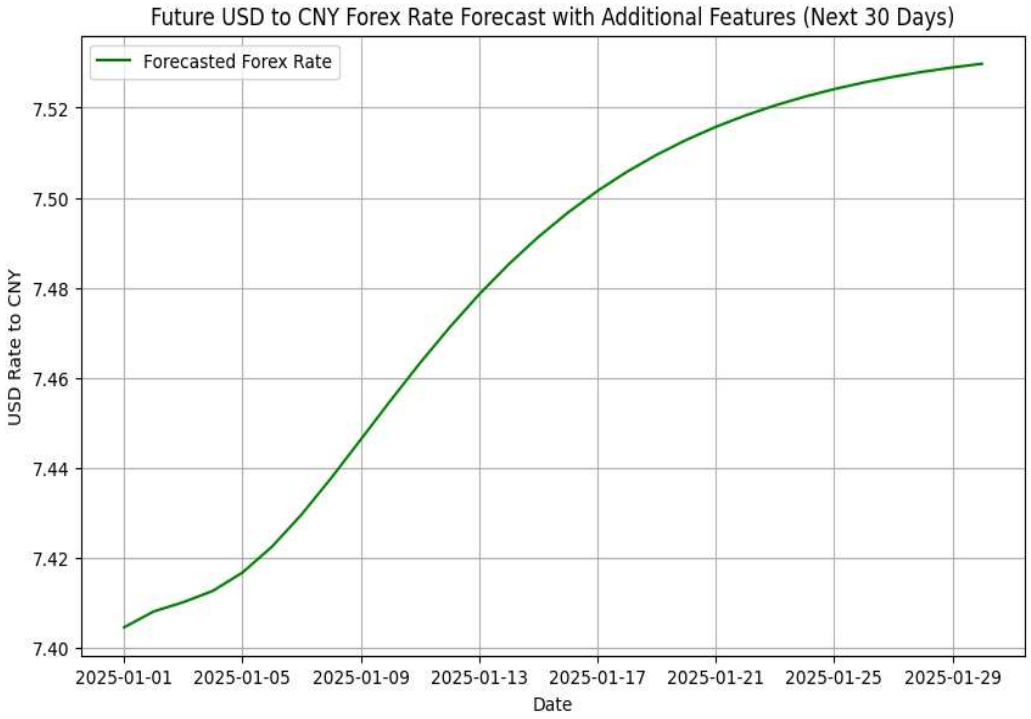


Figure 4.7 forecasting the future pairs USD/CNY rate

CHAPTER 5

Discussion and Future Work

5.1 Discussion

This project aims to analyse the historical data from exchange rate ,CPI ,interest rate and 10-Year bond of China and US.Through the result of the analysis,it can conclude that interest is a major and significant factor which affect the change of the FOREX.Furthermore,CPI and 10-year bond also plays an important role in the movement of the FOREX by directly affecting the interest.

By the use of the LSTM model, it exports the trending of the USD to CNY exchange rate over the next 30 days based on historical data and its relationship with four key macroeconomic factors: CPI for both the US and China, and interest rates and 10-year bond yields from both countries.The LSTM model is good at catching long-term dependencies in sequential data, transforming it to an effective and valuable data for financial time series analysis. It conveys and converts past patterns in the exchange rate and these influencing factors to detect deep relationships and predict future trends.

5.2 Future Work

Due to the limitation of single machine learning,LSTM which is powerful at capturing sequential dependencies,we acknowledge that it may have a shortage in ensuring consistently high accuracy because of challenges such as overfitting, sensitivity of hyperparameters, and the analysis of nonlinear or stochastic relationships in complex datasets.It will lead to misrepresent and slant the final conclusions or findings.

In the future work tackling these inappropriate issues,it is necessary to

concern the possibility of involving hybrid models to facilitate the ability of the current single model,LSTM. Moreover,currently,there are various studies which exhibit kinds of hybrid frameworks such as merging LSTM with ARIMA (AutoRegressive Integrated Moving Average), Monte Carlo Simulation, or HMM(Hidden Markov Model),.According to the research,it is obvious to display the framework can export more accurate and valuable predictions.

REFERENCES

Yuenan Chen1(2023).The Impact of the Interest Rates Raised by Federal Reserve System on the Exchange Rate of the US Dollar Against the Chinese Yuan. .https://www.researchgate.net/publication/376131422_The_Impact_of_the_Interest_Rates_Raised_by_Federal_Reserve_System_on_the_Exchange_Rate_of_the_US_Dollar_Against_the_Chinese_Yuan

Marcin Kolasa, Sahil Ravgotra, and Pawel Zabczyk (2022) .Monetary Policy and Exchange Rate Dynamics in a Behavioral Open Economy Model. https://www.imf.org/-/media/Files/Publications/WP/2022/English

Musandiwa, J., Ngwakwe, C. C. (2023). The Effect of Exchange Rates on Consumer Price Index. Oblik i finansi, 4(102), 47-52. https://doi.org/10.33146/2307-9878-2023-4(102)-47-52

Didarul Islam, Mohammad Abdullah Al Faisal.(2024).Determinants of Uruguay's Real Effective Exchange Rate: A Mundell-Fleming Model Approach. https://arxiv.org/abs/2403.16452

Rosnawintang , Muh. Syarifb, Aini Indrijawatic, Pasrun Adamd, La Ode Saidie .(2019).The Causal Relationship between Exchange Rates and Bond Yield in Indonesia. https://ier.ut.ac.ir/article_81867_7a941a84fd82e8ad3660e821247c333b.pdf

Deniz Can Yıldırım , Ismail Hakkı Toroslu and Ugo Fiore .(2021).https://jfin-swufe.springeropen.com/articles/10.1186/s40854-020-00220-2?

Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," IEEE transactions on neural networks, vol. 5, no. 2, pp. 157–166, 1994.

Ling Qi ,Matloob Khushi ,Josiah Poon .(2020).Event-Driven LSTM For Forex Price Prediction.https://arxiv.org/abs/2102.01499.

Hasna Haifa Zahrah,Siti Sa'adah ,Rita Rismala.(2020).The Foreign Exchange Rate Prediction Using Long-Short Term Memory: A Case Study in COVID-19 Pandemi.https://www.researchgate.net/publication/359168790_The_Foreign_Exchange_Rate_Prediction_Using_Long-

Short_Term_Memory_A_Case_Study_in_COVID-19_Pandemic/citations.