CASCADING MODEL FOR FOREX MARKET FORECASTING USING FUNDAMENTAL AND TECHNICAL INDICATOR DATA BASED ON BERT

Miss Arisara Pornwattanavichai

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science and
Information Technology
Department of Mathematics and Computer Science
FACULTY OF SCIENCE
Chulalongkorn University
Academic Year 2021

แบบจำลองการทำนายอัตราแลกเปลี่ยนเงินตราสากลแบบลำดับขั้นโดยใช้ข้อมูลปัจจัยพื้นฐานและตัวชี้วัดทางเทคนิคบนการเข้ารหัสแบบสองทิศทางโดยอาศัยตัวแปลง

น.ส.อริสรา พรวัฒนวิชัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2564

Thesis Title	CASCADING MODEL FOR FOREX MARKET FORECASTING USING FUNDAMENTAL AND TECHNICAL INDICATOR DATA BASED ON BERT
By	Miss Arisara Pornwattanavichai
Field of Study	Computer Science and Information Technology
Thesis Advisor	SARANYA MANEEROJ
Thesis Co Advisor	SOMJAI BOONSIRI

Accepted by the FACULTY OF SCIENCE, Chulalongkorn University in Partial Fulfillment of the Requirement for the Master of Science

.... Dean of the FACULTY OF SCIENCE

(POLKIT SANGVANICH)

THESIS COMMITTEE

.... Chairman

(MONNAT PONGPANICH)

.... Thesis Advisor

(SARANYA MANEEROJ)

.... Thesis Co-Advisor

(SOMJAI BOONSIRI)

.... External Examiner

(Saichon Jaiyen)

อริสรา พรวัฒนวิชัย : แบบจำลองการทำนายอัตราแลกเปลี่ยนเงินตราสากลแบบลำดับขั้นโดยใช้ข้อมูลปัจจัยพื้นฐานและตัวชี้วัดทางเทคนิคบนการเข้ารหัสแบบสองทิศทางโดยอาศัยตัวแปลง. (CASCADING MODEL FOR FOREX MARKET FORECASTING USING FUNDAMENTAL AND TECHNICAL INDICATOR DATA BASED ON BERT) อ.ที่ปรึกษาหลัก : ศรันญา มณีโรจน์, อ.ที่ปรึกษาร่วม : สมใจ บุญศิริ

ตลาดแลกเปลี่ยนเงินตราต่างประเทศ (Forex) เป็นตลาดทางการเงินที่ใหญ่และมีสภาพคล่องสูงที่สุดในโลก เป็นสถานที่กำหนดอัตราแลกเปลี่ยนระหว่างประเทศ เนื่องจากอัตราแลกเปลี่ยนเงินตราต่างประเทศมีบทบาทสำคัญในเทคโนโลยีด้านการเงินและธุรกิจ จึงทำให้นักวิจัยหลายท่านมีความสนใจในด้านการทำนายอัตราแลกเปลี่ยน โดยลักษณะข้อมูลของ Forex นั้น มีความผันผวนสูง มีคุณสมบัติไม่เป็นเส้นตรง และบางครั้งก็เกิดขึ้นในรูปแบบการเดินสุ่ม (random walk) ทำให้ยากต่อการทำนาย มีงานวิจัยที่เกี่ยวข้องหลายงาน พยายามสร้างการทำนาย Forex โดยการรวมข้อมูลด้านปัจจัยพื้นฐาน (FD) และ ข้อมูลจากตัวชี้วัดทางเทคนิค (TI) โดยข้อมูลจากตัวชี้วัดทางเทคนิคนั้น เป็นสัญญาณที่สะท้อนถึงรูปแบบของราคา ในขณะที่ข้อมูลด้านปัจจัยพื้นฐานจะเป็นข้อมูลที่เป็นตัวชี้วัดสภาพทางเศรษฐกิจของประเทศนั้นๆ อย่างไรก็ตาม เมื่อข้อมูลสองประเภทนี้ถูกนำไปใช้งานจริงกลับมีข้อจำกัดสำคัญถึง 2 ประการ ข้อจำกัดแรกคือปัญหาด้านโมเดล เมื่อใช้แบบจำลองโครงข่ายประสาทเทียมแบบลำดับในการพยากรณ์ ทำให้เกิดปัญหาการสูญหายของค่า gradient (Gradient vanishing problem) และการสูญหายของข้อมูล (information loss) ข้อจำกัดที่สองคือด้านการใช้ข้อมูลปัจจัยพื้นฐาน แม้จะมีผลกระทบต่ออัตราแลกเปลี่ยนเป็นอย่างมาก แต่ข้อมูลจะถูกอัปเดตรายไตรมาส หรือรายเดือน ซึ่งความถี่ในการอัปเดตไม่เท่ากับการเปลี่ยนแปลงของอัตราแลกเปลี่ยนสกุลเงิน ข้อจำกัดนี้มีชื่อว่า ข้อจำกัดการปล่อยของข้อมูลด้านปัจจัยพื้นฐาน ยิ่งไปกว่านั้น โดยปกติในการพยากรณ์อัตราแลกเปลี่ยนที่อาศัยข้อมูลด้านปัจจัยพื้นฐาน และ ข้อมูลจากตัวชี้วัดทางเทคนิค มักถูกนำมารวมเพื่อสร้างการทำนาย โดยให้ความสำคัญเท่ากัน ซึ่งนำมาสู่การพยากรณ์ที่ไม่แม่นยำ เนื่องจากความถี่ของข้อมูลทั้งสองนั้นไม่เท่ากัน ในงานวิจัยนี้ (BERTFOREX) ผู้วิจัยนำเสนอโมเดลการรวมแบบน้ำตก (cascading model) สำหรับการพยากรณ์ราคาอัตราแลกเปลี่ยนโดยอาศัย ข้อมูลด้านปัจจัยพื้นฐาน และ ข้อมูลจากตัวชี้วัดทางเทคนิค บนการข้ารหัสแบบสองทิศทางโดยอาศัยตัวแปลง (BERT) โดยการทำงานในแบบจำลองมีขั้นตอนดังต่อไปนี้ 1) ข้อมูลด้านปัจจัยพื้นฐานจะถูกนำไปสกัดหาลักษณะแฝงของข้อมูล โดยอาศัยการข้ารหัสแบบสองทิศทางโดยอาศัยตัวแปลง 2) เนื่องจากความถี่ของข้อมูลด้านปัจจัยพื้นฐานนั้น เปลี่ยนแปลงช้ากว่าข้อมูลจากตัวชี้วัดทางเทคนิค ลักษณะแฝงที่สกัดได้ของข้อมูลด้านปัจจัยพื้นฐาน จึงถูกรวมเข้ากับ ข้อมูลตัวชี้วัดทางเทคนิค ในรูปแบบน้ำหนักเสริม 3) ข้อมูลการรวมที่ได้จากข้อมูลด้านปัจจัยพื้นฐาน และ ข้อมูลจากตัวชี้วัดทางเทคนิค จะถูกนำไปสกัดลักษณะแฝง โดยอาศัยการข้ารหัสแบบสองทิศทางโดยอาศัยตัวแปลง 4) เพื่อแสดงถึงประสิทธิภาพของโมเดลนี้ รูปแบบการรวมที่สกัดได้ จะถูกนำไปเข้าโครงข่ายประสาทเทียมอย่างง่ายเพื่อสร้างการทำนาย จากผลการทดลอง งานวิจัยนี้สามารถเอาชนะงานวิจัยอื่นๆในเชิง จำนวนสัญญาณที่ถูกต้อง ความไว (sensitivity) ความจำเพาะ (specificity) ความเที่ยงตรง (precision) และค่าทำนายเมื่อผลเป็นลบ (negative predictive value)

สาขาวิชา	วิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ	ลายมือชื่อนิสิต ....
ปีการศึกษา	2564	ลายมือชื่อ อ.ที่ปรึกษาหลัก ....
		ลายมือชื่อ อ.ที่ปรึกษาร่วม ....

6172630023 : MAJOR COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
KEYWORD: Time series forecasting, Deep learning, BERT, Forex

Arisara Pornwattanavichai : CASCADING MODEL FOR FOREX MARKET FORECASTING USING FUNDAMENTAL AND TECHNICAL INDICATOR DATA BASED ON BERT. Advisor: SARANYA MANEEROJ Co-advisor: SOMJAI BOONSIRI

The foreign exchange rate market is the world's biggest and most liquid financial market, and it's where all currency pairs' exchange rates are set. Since foreign exchange (Forex) rates play a critical role in financial technology and business, many researchers are now interested in forecasting them. The characteristics of Forex data, that include fluctuation, non-linearity, and random walk phenomena, make it difficult for forecasting. Several related studies integrate fundamental data (FD) and technical indicator data to generate Forex forecasting signals (TI). TI is a price pattern-based signal, whereas FD is an indicator of the country's economic conditions. Nevertheless, when it comes to deployment, these two indicators have two major drawbacks. Gradient vanishing and information loss occur when modeling a sequential neural network. Furthermore, although FD has a big impact on currency prices, it was updated quarterly or monthly which is not as frequent as price change. This restriction is known as the FD releasing problem. Moreover, Forex forecasting with FD and TI is usually done in equal aggregation, which leads to inaccurate predictions due to unequal data changing frequency. In this paper (BERTFOREX), we introduce a cascading model for forex market forecasting using FD and TI based on BERT (BERTFOREX). The following are the steps in the BERTFOREX processing system: 1) BERT is applied to FD to extract hidden patterns. 2) Because the frequency of FD changes more slowly than that of TI, these hidden FD patterns are aggregated as additional weights for TI. 3) BERT is used to extract the aggregated pattern within TI and FD. 4) The BERTFOREX efficiency is demonstrated by feeding the aggregated pattern into a simple neural network for forecasting. From the experimental results, the proposed method outperforms other methods in terms of correct signal percentage, sensitivity, specificity, precision, and negative predictive value.

Field of Study:	Computer Science and Information Technology	Student's Signature ....
Academic Year:	2021	Advisor's Signature ....
		Co-advisor's Signature ....

ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude to my thesis advisor. Assoc. Prof. Dr. Saranya Maneeroj. Without her assistance, valuable guidance, and continuous encouragement at every step throughout my thesis, I would have never accomplished this work. Next, I feel thankful for my co-advisor. Asst. Prof. Dr. Somjai Boonsiri. She always supports me and keeps tracking me for this thesis. Moreover, I appreciate Dr. Peeraphon Sophatsathit. With his kind heart, he helps me in proofreading my research paper and suggests valuable information. Hence, the publication process could run smoothly with no major errors. In addition, I feel grateful to my family who always supports me and waits for my graduation. Lastly, I would like to thank my responsibility who trying to fight laziness and for always waking me up to do this thesis. Thank you, for every cup of coffee that refreshes me. Lastly, I would like to thank all my learned ML knowledge during my M.S. studies, my hardworking experience with ML, and Dr. Saranya who push and give me a passion to do this thesis and become a data scientist.

Arisara Pornwattanavichai

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

The term "Forex" refers to the foreign exchange rate market, the largest financial market in the world where the daily exchange of massive amounts of currencies takes place. In all major financial centers throughout the world, traders can trade currencies at any time because the forex market is open 24 hours a day and five days a week.

The currency values of a nation against another, commonly known as exchange rates, are determined by the forex market. A straightforward form of foreign exchange involves exchanging one currency for another at a nearby bank. Exchange rates play a critical role in determining the exchange market's dynamics, which are controlled by changes in currency demand and supply over time.

Traders that trade in the forex market can speculate on the direction of the exchange rate. Assume the trader is currently trading on USD/EUR quotations. In that instance, to gain profit on the Forex market, a trader must correctly predict the direction that the exchange rate will go in the future. Assume that the USD/EUR ratio increases when the USD strengthens relative to the EUR, the quote currency. The traders must buy the USD base currency in order to make a profit. However, the USD/EUR depreciates if the US dollar, the base currency, drops. The merchants must sell the basic currency. If not, the traders run the danger of losing money and quitting. As a result, forecasting the direction of the exchange rate is the most crucial task for forex traders to perform.

In today's automated markets, investors can exchange currencies from any place by connecting their PC or mobile phone to the Internet.

Trading currencies is easy, but it is challenging to do it without losing money or making a lot of money. Because the exchange rate is influenced by a number of economic factors, and their interconnections are exceedingly complex. As a result, anticipating changes in foreign currency rates is the most difficult issue in the field of time series forecasting.

Many academics develop forecasting signals by combining fundamental and technical investigations with various trading techniques. Fundamental analysis is concerned with the monetary, social, and political factors that influence price direction. These variables are referred to as fundamental data.

To produce sell and purchase signals, technical analysis primarily employs price information such as closing price, open price, high price and low price. However, mathematical price modifications, known as technical indicators, are sometimes included.

In this work, fundamental data (FD) and technical indicator data (TI) used in a fundamental and a technical analysis technique are concentrated to forecast the direction of price in the future.

In most cases, either FD or TI is inputted to the model to generate forecasting.

Predictions are produced in [1] and [2] using FD as an input and the extracted correlation between price and economic data. The model structure chosen is the Multilayer Perceptron (MLP). Even though FD-based model can provide forecasting signals, the results are unreliable because the FD is rarely updated. Since FD is only delivered on a set day every week, quarter, or month, it is thought to be of no lasting significance. On the other hand, exchange rates are continually changing and evolving over time. The model will therefore have trouble capturing the impact of the FD-currency pricing relationship.

On the other hand, TI is used as the model input by Yao and Tan [3]. In order to produce predictions, moving average (MA) intervals of various durations are applied to neural networks (NN). The unconsidered influence from FD, notwithstanding the possibility that it signals a price reversal, is a drawback of the TI-based method. The TI model nevertheless gains from rapid information updates and quantitative pricing in a number of essential qualities for unpredictable price forecasts.

Several recent research have advocated combining FD and TI to produce forecasting. AbuHamad et al. [4] calculated FD and TI using independent MLP models. The predicted outputs of sub-model are then merged in knowledge based rules to produce the final output.

Yldrm et al. [5] used TI and FD on Long-short term memory (LSTM) individually. Their model aggregated LSTMs results using a set of rules. The final forecast is then created. As humans create the integration rules, these restrictions might reduce the model's ability to forecast.

The drawbacks of existing models are outlined below. Even though sequential neural networks are not only well-known but also widely used in exchange rate prediction, some model limitations such as gradient vanishing difficulties and information loss are remain due to extended input sequences in Recurrent Neural Networks (RNN) and forget gate mechanism in LSTM.

As a result, our model employs Bidirectional Encoder Representations from Transformers (BERT), that include a self-attention mechanism. The self-attention mechanism allows our model to extract and remember the relationship between the current and other date, irrespective to their distance. BERT is often utilized in Natural Language Processing to handle text processing. Nevertheless, both time series and text are sequential data. In this work, BERT is utilized to include and use critical information from the days preceding and following each date (context). The benefits of BERT keep our study from losing critical data and improve the accuracy of exchange rate forecasting.

However, there is a major issue remains which is FD absent value. Because FD is rarely updated, it may be regarded as a missing data at unreleasing dates, despite the fact that currency prices change on a regular basis. As a result, the model will struggle to identify modest changes that may have a significant influence on model prediction. As a result, the more frequently FD is updated, the better predictive accuracy is produced. The frequency of change for TI and price, however, is same. Considering that when we calculate TI using a mathematical equation. We might enter the past price with a predetermined frequency. The same frequency TI will then be produced. In our suggested cascade aggregation approach using BERT, named Cascading Model for Forex Market Forecasting Using Fundamental and Technical Indicator Data Based on BERT, we aggregate FD to TI to address the FD problem (BERTFOREX).

We must use the proper data for the representation of turbulent price movement if we want to respond swiftly to frequent Forex change. As a result, TI is selected as the main feature of BERTFOREX, with FD adding weight to TI throughout the aggregation phase.

We choose to apply BERT to discover the underlying relation among a quarter of the FD before starting the aggregation process due to the FD absence value problem. This results in our presented cascade model for forecasting forex.

The cascading design enables this work to organize incoming information appropriately. Initial, using the first BERT, FD is modified as an additional weight for TI data (BERT1). Next, the extra weight from FD that was retrieved is added to TI. The latent characteristic of the integrated representation is obtained using the second BERT to guarantee that the result contains all pertinent data from the date that affects the target date (BERT2). The result of this step equates frequency change to currency price change by representing TI and FD depending on other affected dates. The cascade aggregate frees BERTFOREX from the FD releasing challenge and generates an accurate prediction. In BERTFOREX, we show the effectiveness of our aggregation pattern from cascaded BERTs. Lastly, we use a simple neural network to show the effectiveness of our aggregate pattern from cascaded BERTs in BERTFOREX prediction.

CHAPTER 2 RELATED WORKS

In addition, many research illustrate that applying both fundamental and technical data provide better forecasting results than predictions with either of them. There are 2 types of forex forecasting model as follows:

1. Fundamental data (FD): the fundamental data can illustrate the economic condition of the target currency country which affects to its currency exchange rate.

2.Technical indicator data (TI): technical indicator data is calculated from the historical of currency data. Hence, it changes according to the price action all the time, exposed the traders' trading behavior and currency price pattern.

The two primary types of all prior Forex forecasting models [1-4], [5-9] are as follows: 1.) Either FD or TI based prediction method 2.) FD and TI aggregation based prediction method

2.1 EITHER FD OR TI BASED PREDICTION METHOD

Either fundamental data or technical indicator data is processed as input of the model to make forex price prediction. There are 3 previous research perform this model type as follows:

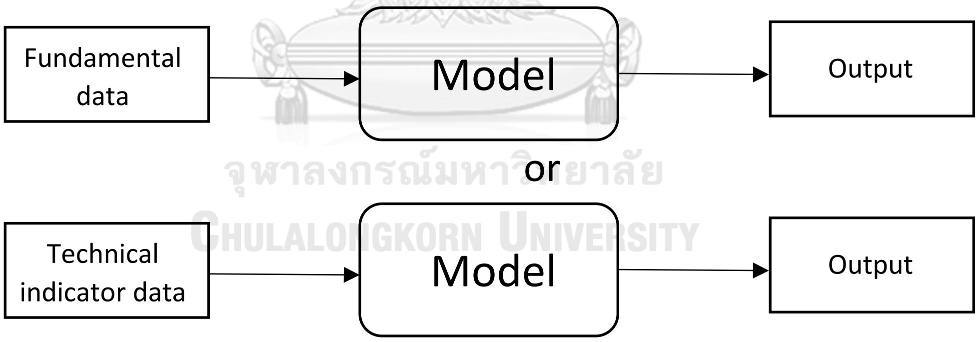

Figure 1. Either FD or TI based prediction method

2.1.1 NONLINEAR PREDICTION OF EXCHANGE RATES WITH MONETARY FUNDAMENTALS

This paper [1] used a Three Layer feed forward Neural Network (Multilayer perceptron) to forecast monthly closing prices of USD/JPY, USD/DM, USD/GBP, and USD/CAD using fundamental data such as money supply by M1, real income by industrial production in each of the countries, interest rate: treasury bill rates for Britain, Canada, and the United States, and money rates as an alternative measure of interest rate for Japan and Germany.

To measure the ability of the model, the model was compared with other 3 models as follows: Linear monetary model (LR), Random Walk model (RW) and NN model without fundamentals.

They discover the lack of ability of normal NN in forecasting closing price movement. Their model produced higher RMSE than random walk model, simple linear monetary model and NN without fundamentals in all currency pairs especially when the forecast horizon is lengthens. From the result, the researchers refer to another research report that the forecasting ability of linear model improves as forecast horizon lengthen, the pattern is not discovered within NN framework.

2.1.2 A CASE STUDY ON USING NEURAL NETWORKS TO PERFORM TECHNICAL FORECASTING OF FOREX

This paper [3] proposed using only technical indicator data such as historical price change, MA5, MA10, MA20, and MA60 to capture the underlying rules of movement in currency price change, where MA stands for Moving Average value of a specific number of days in the past. The output of the model is weekly closing price of 5 currency pairs as follows: USD/AUD, USD/CHF, USD/DEM, USD/GBP and USD/JPY. And the prediction closing price are applied with 2 different trading strategies to make decision buy/sell where x_t denote the actual level at time t, $\hat{x}_t$ denote the prediction of neural networks. These two different strategies are selected based on the prediction error. If the sign of next date prediction error (x_t - $\hat{x}_t$) is not similar to the sign of current date prediction error (x_{t+1} - $\hat{x}_{t+1}$), strategy1 is selected. Otherwise strategy2 is selected.

Trading Strategy 1:

$$\text{If } (\hat{x}_{t+1} - \hat{x}_t) > 0 \text{ then } \textit{buy} \text{ else } \textit{sell}. \quad (1)$$

Trading Strategy 2:

$$\text{If } (\hat{x}_{t+1} - x_t) > 0 \text{ then } \textit{buy} \text{ else } \textit{sell}. \quad (2)$$

The result showed that NN with only simple technical indicator data can make useful forecasting and gain significant profit for out-of-sample data without any extensive market data. But it is not easy to make profits with those data because there are several factors affect to the prediction such as frequency of sampling, choice of network architecture, forecasting periods.

Moreover, there are some interesting observations from the result. First, Yen is hard to forecast more than other currencies because the trader of Yen market act quickly when any sign of market appears. Therefore, NN with only technical indicator data is not enough to make Yen prediction. Second, trading strategy 2 is more powerful in using NN forecasting outcome. Lastly, there need to have efficient ways to measure the performance of the price prediction model not only NMSE but also other method such as gradient and profits.

2.1.3 FORECAST FOREX WITH ANN USING FUNDAMENTAL DATA

This paper [2] proposed using fundamental data and a multilayer perceptron (MLP) to forecast forex prices. Because the target currency pair is derived from a financial relationship between two countries, the United States and Europe, the researcher assumed that the factors influencing currency price movement are those of these two countries. Therefore, EUR and USD fundamental data are chosen as follows: CPI, GDP, TS (Trade Surplus), IR (Interest rates).

The fundamental data normally is updated on a quarter basis, but currency price is in daily form. So, there are two approaches to overcome this different time frames issue. First, the fundamental data is kept as constant and fed to model as daily input. Second, linear or polynomial method is used to interpolate the data between two quarters. But the experiments showed that two approaches do not have significant different improvement between them.

Although, the fundamental data is important to exchange rate movement in real world, but ANN cannot capture underlying relationship between them. So, they cannot improve ANN predictive performance. This limitation maybe caused of the frequency of fundamental data releasing, called fundamental data releasing problem. If they are more frequent updated, the predictive performance of ANN may perform better.

Recently, there have been attempts to combine FD and TI, leading to a novel data representation. It depicts the state of the economy in the nation as well as current pricing movements. According to numerous research, forecasting based on this aggregate yields reliable Forex forecasting outcomes.

Based on previous research, the aggregation model uses sequential neural networks to predict data from both FD and TI combined together. Compared to the model that used either one, the outputs that were generated were more precise.

Several Forex forecasting methods combine both FD and TI to produce forecasts using heuristic criteria.

2.2 FD AND TI AGGREGATION BASED PREDICTION METHOD

Fundamental data and technical indicator data are learned separately in two sub-models. The results of these two models are then aggregated using heuristic rules [4, 10] [5] to determine whether the market will be bullish or bearish.

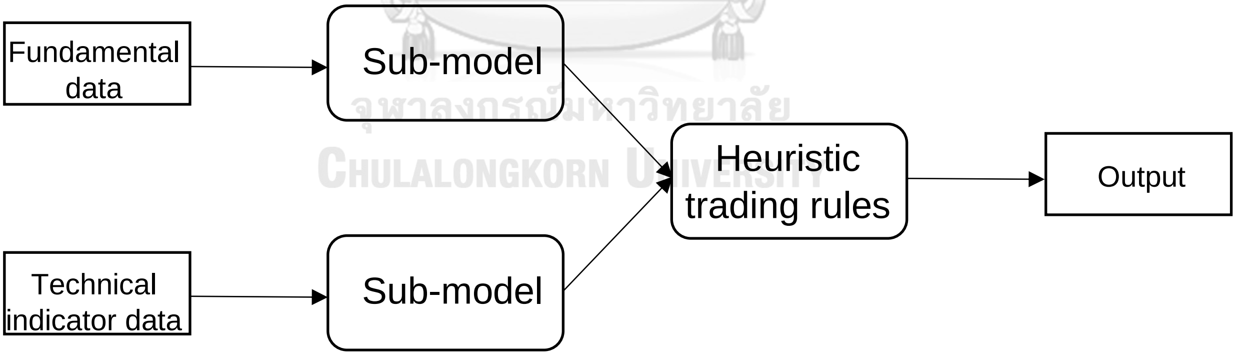

Figure 2. FD and TI aggregation based prediction method

2.2.1 EVENT-DRIVEN BUSINESS INTELLIGENCE APPROACH FOR REAL-TIME INTEGRATION OF TECHNICAL AND FUNDAMENTAL ANALYSIS IN FOREX MARKET

Both currency data and fundamental data, such as currency close price and nine US macroeconomic data (CPI, GDP, Housing starts, Initial unemployment claims, Nonfarm payrolls, PPI, Retail sales, Trade Balance, Unemployment), are applied to two different MLP sub-models in their work [4].

Currency data is applied to the first sub-model while the second one applies fundamental data. Signal output from two sub-models were aggregated to the final model using their manually defined knowledge-based rule to generate signal BUY or SELL of CAD/USD, CHF/USD and USD/EUR in eleven timeframes.

Knowledge-based rule

RULE I

If exchange-rate_signal is equal macroeconomic_signal (3)

Then signal adopted.

RULE II

If exchange-rate_signal differs from macroeconomic_signal (4)

Then signal aborted.

The result showed that the higher frequency, the lower accuracy because the number of observed data in high frequency is higher than the number of observed data in low frequency. As the irregularity nature of financial data appears more obvious in high frequency data, makes it harder to captures in any well-known method. The aggregated model can predict the number of correct signals about 84-91% in CADUSD, 86-93% in CHFUSD and 71-89% in USDEUR especially in timeframe 1-Day minutes and 1-Week.

2.2.2 FORECASTING DIRECTIONAL MOVEMENT OF FOREX DATA USING LSTM WITH TECHNICAL AND MACROECONOMIC INDICATORS

A hybrid model based on Long Short-Term Memory was proposed (LSTM) [5]. To generate the two sub-model signal outputs, two LSTM models are run separately with fundamental data and technical indicator data. For fundamental data, there are 7 kinds of them are applied in the first sub-model which are Close S&P500, Close DAX, Interest rate GER, Interest rate EU, FED Funds Rate, Inflation Rate EU and Inflation Rate USA. While, the second sub-model used MA, MACD, ROC, Momentum, RSI, BB, CCI technical indicator data as its input feature set. In order to determine the EUR/USD close price forecast in the form of BUY/SELL/NO ACTION, the findings from the earlier sub-models are then combined using their manually established set of rules.

To compare the accuracy of the hybrid model, there are 4 types of models were chosen which are ME_LSTM (Macroeconomic data with LSTM), TI_LSTM (Technical Indicator data with LSTM), ME_TI_LSTM (Macroeconomic data and Technical Indicator data with LSTM) and Hybrid model (Aggregated Macroeconomic data and Technical Indicator data with set of rules based on LSTM).

The experimental can be categorized into 3 groups up which are forecast 1 day ahead, 3 days ahead and 5 days ahead, respectively. From the results of 3 forecasting timeframes, individual model gets slightly better results than ME_TI_LSTM, but hybrid model provides the best profit accuracy.

In a brief, heuristic rules are manually established by persons as part of the aggregation process. The model's forecasting accuracy is crucially constrained by these heuristic rules. In the sections that follow, we will describe them in depth and propose alternatives.

MODEL LIMITATION

Forex forecasting model always utilizes RNN or LSTM [11-16]. From characteristic of LSTM, the information flow is controlled by gates and historical data from previous state is passed through the next state by cell state. Hence, gradient vanishing problem that occurred in RNN is eliminate. But there still has some limitation of LSTM, when the sequence of data is too long, some critical data in the past can be lost. LSTM does not keep all the data in the past. Rather, it is merely a decision that which historical data should be kept for a period of time.

Memory constraints were always present in a conventional sequential neural network. The impact of historical price on current price is reduced as the length of the information sequence is increased. Even if the predicted price is similar to the pattern from two years ago data, the impact of data from two years ago is less than data from the last two days before the prediction. A self-attention mechanism is applied to overcome memory constraints. Self-attention is typically used in Transformer [9] to improve machine translation performance. Self-attention enables the model to understand how the present word in the sequence relates to the ones that come after it. No matter how far apart the words are, each word will preserve the hidden representation of every other word in the input sequence since it will be represented by a score obtained from prior words.

Nevertheless, focusing only on one component of price is insufficient. Multi-head attention is the process of using self-attention repeatedly in parallel to gain additional price views [17]. Using numerous heads of attention, the method can aggregate the knowledge value from different self-attentions. Consider that there are five distinct categories of head attention. Heads 1 through 5 gain knowledge of price movement patterns, quantity, fluctuation, correlation, and trend. The result of combining and projecting these five bits of data is a useful representation.

In general, many language models make use of the capability of multi-head attention to construct an embedding of every word in text input. The most popular language model is BERT [18]. A representation for each word location is learned and obtained by BERT using a bidirectional learning technique that takes into account both left and right contexts. Compared to other machine translation, this will produce more accurate findings.

DATA LIMITATION

FD is frequently used when talking about data because it has a significant effect on the price when information is published. As mentioned before, FD and TI are typically coupled to produce forecasting signals. The output signal could, nevertheless, be erroneous because of FD constraints, which is termed as the FD releasing problem. Furthermore, as FD is normally updated every three months, it is characterized as a loss of value over time (between two successive changes) despite changing exchange rates. As a result, the model will have trouble capturing the effects of the link between FD and currency prices, especially in the FD-based prediction model. As a result, FD is changed more often the greater the predictive performance.

While the price signal is equal to the change in TI's frequency. The MA (n) technical indicator, for instance, computes the MA value using the most recent (n) data points, including the current time step, and is used to visualize price patterns. This calculation causes the MA value to fluctuate each time a new data point is entered.

In order to allay these problems, BERTFOREX incorporates both TI and FD as model inputs and creates a novel cascade aggregation method that enables the algorithm to take into account various TI and FD frequencies.

CHAPTER 3 PROPOSED METHOD

The presented BERTFOREX cascade model is shown in Fig. 3. It consists of three parts: foundational BERT data, aggregate TI to FD, and prediction procedure, where the output of one part serves as the input for the next.

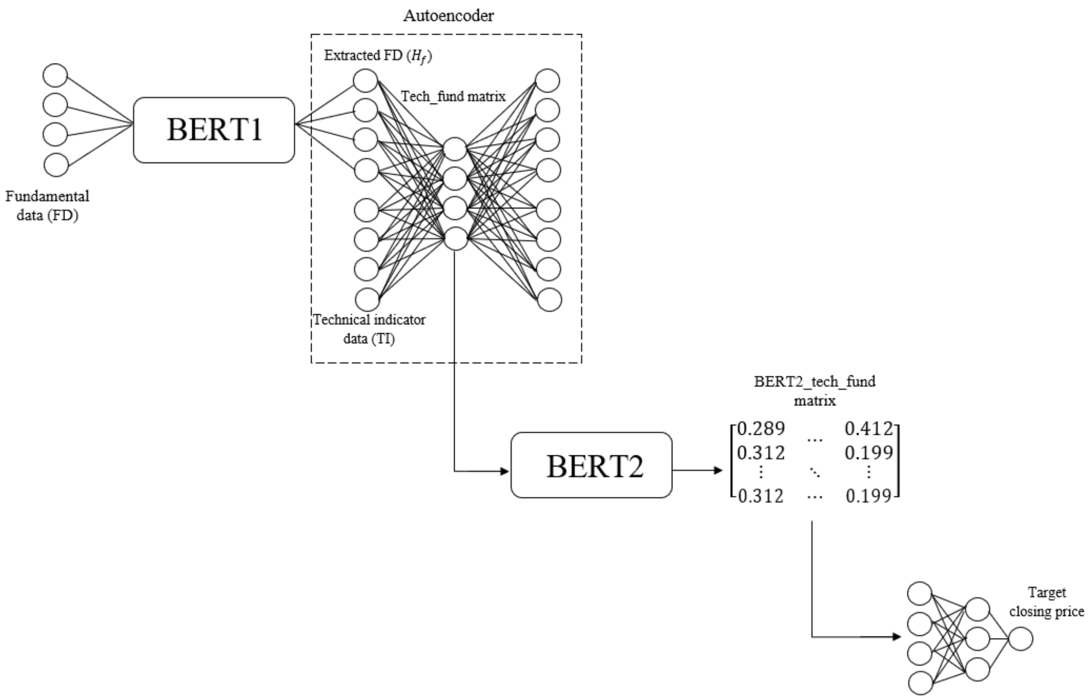

Figure 3. BERTFOREX overview

First, a set of FDs is fed into the *BERT1* process, which results in the extraction of the basic representation matrix (H_f). The latent pattern of FD based on its context is denoted by H_f. Because we plan to give closing price forecasting primarily for TI, which directly follows the currency price, H_f is treated as an additional weight of TI when both TI and H_f are collected on the same day. Because TI and H_f have different meanings and are located in different dimensions, we use an autoencoder to map TI to H_f.

An autoencoder, in general, is a feature extraction method that decreases the dimension of input data through compression and regeneration. In order to reconstruct the input as accurately as possible, the autoencoder hiding layer combines both the important features from the input layer and filters out the noise. We choose the value of the autoencoder hidden layer as the underlying representation between TI and H_f, known as the *tech_fund* matrix, as the input to BERT2. Then, BERT2 analyzes the *tech_fund* matrix to produce the *BERT2_tech_fund* matrix, where *BERT2_tech_fund* reflects the underlying pattern of TI and FD based on their context. The results are fed into an ANN, which forecasts the closing prices. All model procedures are depicted graphically in Fig. 3.

3.1 FUNDAMENTAL DATA ON BERT

BERTFOREX processes FD in *BERT1* to extract a latent representation of FD depending on its context to avoid the occurrence of FD absence value. As indicated in Table 1, BERTFOREX employs 14 different types of FD. These are picked based on previous research and their influence on the target currency pair, such as USD/EUR. The DAX, FFD, GDP, Housing Start Price, Inflation Rate EU, Initial Unemployment Claim, Interest Rate EU, Nonfarm Payrolls, PPI, Retail Sales, SP500, Trade Balance, Unemployment Rate, and CPI are the indicators.

Data release	Source	Frequency
U.S. FFD	Macro trends	Daily
U.S. GDP	BEA	Quarterly
U.S. Housing Start Price	Census	Monthly
U.S. Initial Unemployment Claim	ETA	Weekly
U.S. Nonfarm payrolls	BLS	Monthly
U.S. PPI	BLS	Monthly
U.S. Retail Sales	Census	Monthly
U.S. Trade Balance	BEA	Monthly
U.S. Unemployment rate	BLS	Monthly
U.S. CPI	BLS	Monthly
DAX	Macro trends	Daily
S&P500	Yahoo finance	Daily
Inflation Rate EU	Euro-area-statistics	Monthly
Interest Rate EU	Euro-area-statistics	Monthly

Table 1. Elementary information of fundamental data

The value of each FD in Table 1 is in a particular range, that has a direct impact on the model learning performance. Before initiating any machine learning procedure, FD is initially modified at the start of the BERTFOREX procedure by normalizing the range of each data type into zero to one. In most cases, the data point and its representation in the *BERT1* input sequence are a word and its embedding.

However, our model used a day and daily FD instead. Furthermore, the *BERT1* vocabulary contains unique terms, each of which corresponds to a day represented by FD. As a result, the total number of vocabularies in our dataset equals the entire number of days. Figure 4. *BERT1* depicts the structure of FD processing on *BERT1*. The pre-training step in *BERT1* is used to uncover a new FD latent pattern that happens depending on its context vector (h_{ft}) of each day that is represented by its context or other influencing days.

The following is a description of the pre-training step in *BERT1*. First, FD is used as a word embedding in *BERT1* to represent a date. Our model then randomly rewrite as a [MASK] token eventually. The model attempts to forecast the masking day based on its context, with each day's output represented by a final hidden vector (h_{fn}). This h_{fn} vector is converted to a log_probs vector by using a linear layer and a softmax layer. The model then calculates the index of log_probs with the highest probability. This index is related to the day-in-day matrix index. This day-in-day matrix will be referenced in the final concealed vector.

No.	Technical indicator data
1	Moving Average (MA)
2	Exponential Moving Average (EMA)
3	Moving Average Divergence Convergence (MACD)
4	Single Line
5	Relative Strength Index (RSI)
6	Middle Boillinger Band
7	Upper Boillinger Band
8	Lower Boillinger Band

Table 2. Elementary information of technical indicator data

The prediction loss is calculated by subtracting the one-hot encoding of the actual day from the anticipated day from log_probs. The model uses this loss to begin the backpropagation training process, adjusting the weight until the loss is reduced. *BERT1* is trained until the prediction loss is as small as possible. All output vectors h_{fn} are stored in H_f. H_f has the shape mxn, where m is the size of the input embedding and n is the number of days in the dataset. The *BERT1* output is H_f, which describes the hidden pattern of FD depending on its context.

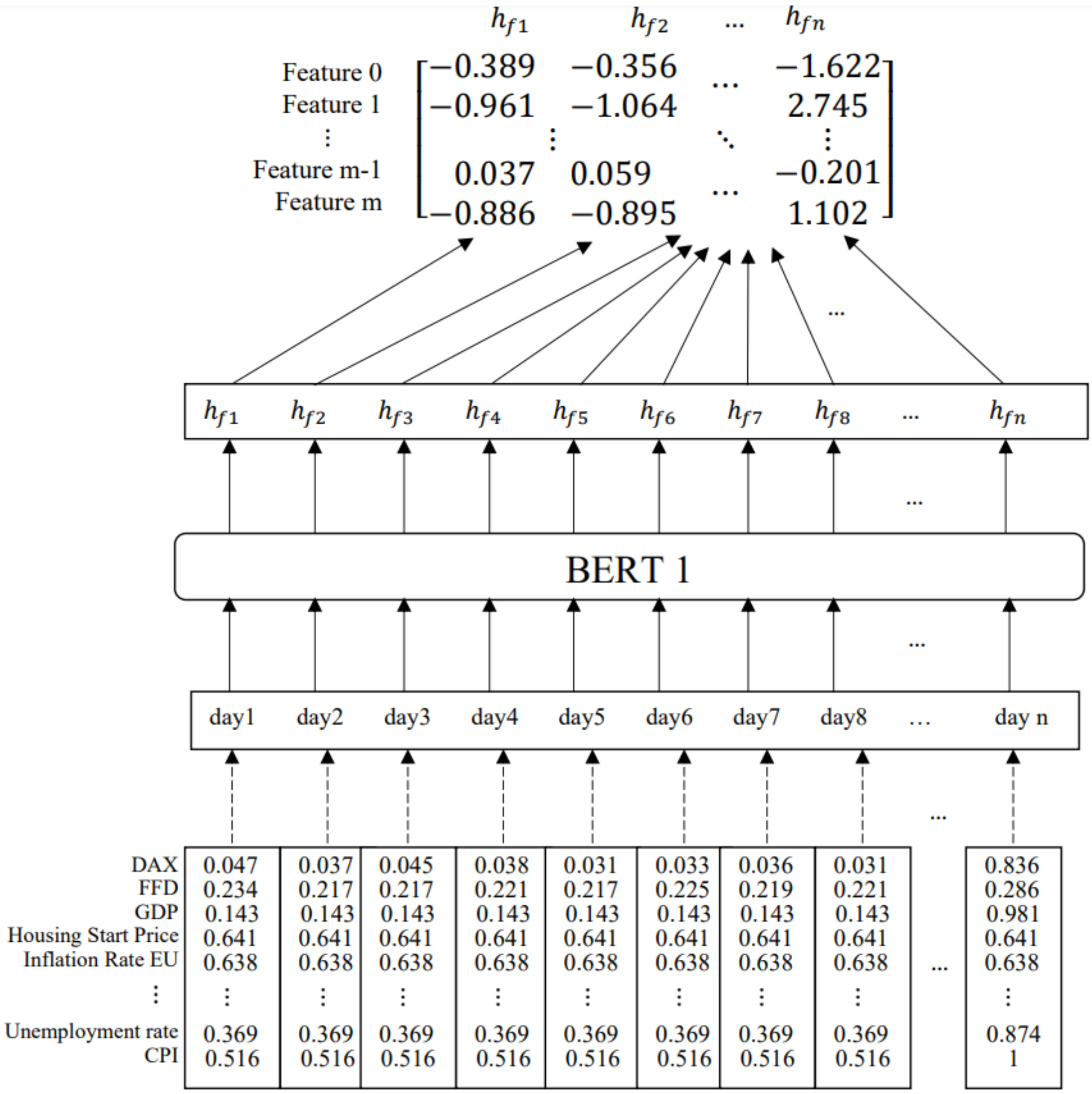

Figure 4. Structure of BERT1

3.2 AGGREGATE TECHNICAL INDICATOR DATA WITH FUNDAMENTAL DATA

A new representation of FD and TI is discovered in this process by integrating FD outcome from *BERT1* with TI utilizing autoencoder and *BERT2*.

Because TI and FD represent information in various ways, their dimensions are put to different dimensional spaces as well. As a result, we cannot immediately aggregate them. As a result, as shown in Fig. 3, the autoencoder is used to map FD and TI to the same dimensional embedding space.

When the training process is finished, the autoencoder hidden node values are chosen as the process output, as indicated by the *tech_fund* matrix.

This *tech_fund* matrix represents TI and FD in the same dimensional space, while all input arrangement is still the same. The final hidden pattern extraction procedure then continues by utilizing BERT a second time, referred to as *BERT2*. Similarly to *BERT1*, the *tech_fund* matrix is input into *BERT2* to uncover integrated hidden representation, particularly on its left and side side.

We train *BERT2* till the model loss is as small as possible. The integrated hidden representation outcome is then built, denoted by the *BERT2_tech_fund* matrix in Fig.3 . Based on consecutive days and appropriately forecasted days, the *BERT2_tech_fund* matrix represents both TI and FD.

3.3 PREDICTION TASK

At this step, rather than another sequential neural network, we predict Forex using a three-layer feedforward neural network.

We want to demonstrate that a three-layer feed forward with meaningful data representation outperforms a sequential neural network with non-meaningful data representation in terms of prediction accuracy.

At time step *t*, we feed the *BERT2_tech_fund* matrix into the ANN of the BERTFOREX forecasting task. The target closing price at time step *t+1* is the output. This is known as rolling windows [19]. In this experiment, the size of the rolling windows and the forecasting horizon are set to 1.

CHAPTER 4 EVALUATION

To indicate the accuracy of BERTFOREX, we compare it to three previous researches that use either FD or TI or both but aggregate them using heuristic methods. Moreover, in the last two tests, we compare the capability of BERT in hidden relation extraction for exchange rate forecasting to that of sequential neural networks.

The next sections will go over the experimental details and data preparation, baseline, settings, evaluation measures, and findings.

4.1 DATA PREPARATION

Our experiment's input data are shown in Tables1, Table2. Between 03/February/2003 and 28/February/2020, FD and TI were collected. The dataset spanned 4,455 days. Because we intended to build forecasts with the economies of the United States and Europe, we chose USD/EUR as our target currency pair and set the data frequency to daily. The dataset is divided into groups to cascade the settings of each component, as shown in Table1, Table2.

4.2 BASELINES

To indicate the efficacy of the BERTFOREX aggregation procedure, we compared its efficiency to that of three other Forex forecasting models: the FD based prediction model (FD model) [20], the TI based prediction model (TI model) [21], and the FD and TI aggregation based prediction method that used heuristic rules [5]. To eliminate bias, these three tests were run on the same dataset. Table 3 depicts all of the method's properties.

Methods	Fundamental data	Technical indicator data	Aggregation
FD model	✓	×	×
TI model	×	✓	×
FD and TI model-based prediction	✓	✓	Heuristic aggregation rules
BERTFOREX	✓	✓	Automatically aggregate
FD+BERT1+ANN	✓		×
TI+BERT1+ANN	v	✓	×

Table 3. Comparison of the properties of all techniques.

Furthermore, we demonstrated BERT's effectiveness in hidden relation extraction for exchange rate prediction versus the sequential neural network by running two more models and comparing them to LSTMs [20, 21]. Those two models are TI+*BERT1*+ANN and FD+*BERT1*+ANN, with *BERT1* being a BERTFOREX module that extracts latent relationships from input data.

We compare TI+*BERT1*+ANN and FD+*BERT1*+ANN to the TI model (TI+LSTM) and the FD model (FD+LSTM). The following is a summary of each approach detail:

1. TI model

[21] utilize LSTM for exchange rate forecasting based TI. The model parameters are adjusted to find the best parameters for the lowest forecasting error.

2. FD model

[20] applies LSTM to FD to make forecasts since the researcher believes that the rate is influenced by a variety of economic factors.

3. FD and TI model-based prediction

To generate the two sub-model signal outputs, two LSTM models [5] are run separately with FD and TI. To obtain the closing price, the results of previous sub-models are integrated using their manually established set of rules.

4. TI+BERT1+ANN

This model extracts the latent relationship within TI using BERT1. The extracted TI relationship is then fed into ANN to generate the forecast.

5. FD+BERT1+ANN

The FD hidden relationship is extracted from FD using *BERT1*. The extracted FD is then transferred into ANN to create a currency price forecast.

4.3 EXPERIMENTAL SETTINGS

BERTFOREX is composed of three cascade components with varying experimental parameters. Following is a more detailed illustration of each component.

1. *BERT1* and *BERT2* settings

In the BERT learning process, the input is typically sequential data with a maximum sequence length *t*. The dataset includes a list of currency prices organized by date. As illustrated in Fig. 5, we divided the dataset into several sub-sequences, with each sub-sequence formed by separating every max sequence length *t*. When *n* sub-sequences are generated, each sub-sequence consists of three parts: a training set, a validation set, and a test set. The training set is *1* to *t-2* points long, whereas the test and validation sets are *t* and *t* points long, respectively. An embedding vector of day *t* in *BERT1* is represented by an FD of *t*, but an embedding vector of day *t* in *BERT2* is represented by a day *t* vector from the *tech_fund* matrix.

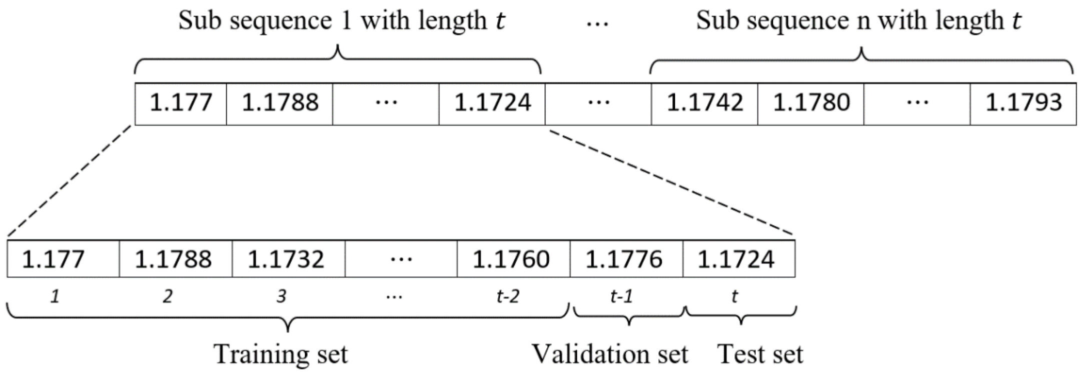

Figure 5. BERT1 and BERT2 settings

2. Autoencoder settings

Using an autoencoder, this procedure maps H_f and TI to the same dimensional space. H_f and TI are two-dimensional matrices with the dimensions $totalFeaturesH_f \ x \ m$ and $totalFeaturesTI \ x \ m$, respectively, where *m* signifies the total number of days. H_f and TI are divided into two parts: training and testing sets. The training set contains 80% of the data, whereas the test set contains 20%. The training set is then supplied into an autoencoder, which is then trained until the test set loss is reduced. Finally, all H_f and TI values are fed into a trained autoencoder. The hidden layer values were chosen as a *tech_fund* matrix.

3. Prediction settings

In the prediction phase, we use a financial cross-validation approach called as purging. Purging is used in BERTFOREX to prevent data leaks by guaranteeing that predictions on the test set are devoid of look-ahead bias from the training set. The dataset is split into three parts: training, purging, and testing. The training and test sets include 80% and 20% of the data, respectively, while the purge length is equal to TI's maximum calculation length. For example, in our model, if we use MA (10), EMA (12), and EMA (50) as TI, the longest computation period is 50. As a result, the purge length will be set to 50. To purge data, we locate the purging period's center on the boundary between the training and test sets. For example, if the dataset is 100 days long, the training set begins on day 1 and ends on day 80, whereas the test set begins on day 81 and ends on day 100. If the purge length = 10, the days 75 to 80 and 81 to 85 are removed from the dataset.

Before evaluating the models' binary classification performance, we convert the prediction results to numerical values using Equation 3-6 to determine the binary classification value in 0 or 1.

$$binary\_actual_t = \begin{cases} 0 \text{ if } Actual\ signal_t < 0 \\ 1\ Otherwise \end{cases} \tag{3}$$

$$binary\_predict_t = \begin{cases} 0 \text{ if } Predict\ signal_t < 0 \\ 1\ Otherwise \end{cases} \tag{4}$$

$$list\_actual\_binary = [binary\_actual_0, \ldots, binary\_actual_t] \tag{5}$$

$$list\_predict\_binary = [binary\_predict_0, \ldots, binary\_predict_t] \tag{6}$$

4.4 EVALUATION METRICS

The experimental results are measured in two different ways to demonstrate the results of classification performance: direction measurement and classification accuracy.

1. **Direction measurement**

We use a percentage of correct direction signals as a metric to assess the model's efficiency in terms of direction. The percentage of correct direction signals is high if the model has good predictive accuracy, as shown by Equations 7-12

$$Actual\ signal_t = actual\_price_{t+1} - actual\_price_t \tag{7}$$

$$Predicted\ signal_t = pred\_price_{t+1} - pred\_price_t \tag{8}$$

$$if = \begin{cases} correct\_up & \text{if } Actual\ signal_t \geq 0\ and\ Predict\ signal_t \geq 0 \\ incorrect\_up & \text{if } Actual\ signal_t \geq 0\ and\ Predict\ signal_t < 0 \end{cases} \tag{9}$$

$$if = \begin{cases} correct\_down & \text{if } Actual\ signal_t < 0\ and\ Predict\ signal_t < 0 \\ incorrect\_down & \text{if } Actual\ signal_t < 0\ and\ Predict\ signal_t \geq 0 \end{cases} \tag{10}$$

$$Total\_correct\_signal = \sum_{i=t}^{n} correct\_up_t + correct\_down_t \tag{11}$$

$$\%correctdirectionsignals = \frac{Total\_correct\_signal * 100}{Totalsignal} \tag{12}$$

where $actual\_price_t$ is the actual closing price of day *t* and $pred\_price_t$ is the predicted closing price of day *t*, which is BERTFOREX's output. The closing price change from day *t* to *t + 1* is represented by $Actual\ signal_t$ and $Predict\ signal_t$. Using the conditions in Equations 7 - 8, we calculate the number of $correct\_up$, $incorrect\_up$, $correct\_down$, and $incorrect\_down$. If $Actual\ signal_t$ and $Predict\ signal_t$ are both higher than 0, the number of $correct\_up$ increases by one, while the number of $incorrect\_up$, $correct\_down$, and $incorrect\_down$ remains in place.

The percentage of correct direction signals in the scalar value gives us the overall correct prediction results. However, this metric is insufficient to determine how good the model's classifier is. If the target dataset is highly imbalanced, with 80% of the data being decreasing signals, the model will predict that all signals will be decreasing signals. The correct percentage of the resulting signal is 80%, which indicates a high level of prediction accuracy. In fact, our model is ineffective because it is unable to predict any increasing signal.

As a result, supplementary classification evaluation metrics, such as sensitivity, recall, specificity, precision, and negative predictive value, are introduced to recompense for the shortcoming, as expressed by Equations 13-16.

2. Classification accuracy

Sensitivity value: The proportion of positive items that accurately predict among all positive items in the dataset is termed to as the sensitivity value. For instance, the proportion of correctly predicted rising signals among all rising signals in the dataset is provided by

$$Sensitivity = \frac{TP}{TP+FN} \tag{13}$$

Specificity value: The proportion of negative items that accurately predict among all negative items in the dataset is termed to as the specificity value. For example, among all declining signals in the dataset, the proportion of declining signals that are accurately predicted is given by

$$Specificity = \frac{TN}{TN+FP} \tag{14}$$

Precision value: The proportion of positive items that accurately predict among all predicted positive items is termed to as precision value. e.g., the proportion of correctly predicted rising signals among all predicted rising signals is provided by

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

Negative Predictive Value (NPV): The proportion of negative items that accurately predict among all predicted negative items is termed to as the Negative Predictive Value (NPV). For instance, the proportion of accurately predicted decreasing signals among all forecasted decreasing signals is provided by

$$NPV = \frac{TN}{FN+TN} \tag{16}$$

True positive, false positive, true negative, and false negative are symbolized by *TP*, *FP*, *TN*, and *FN*, respectively.

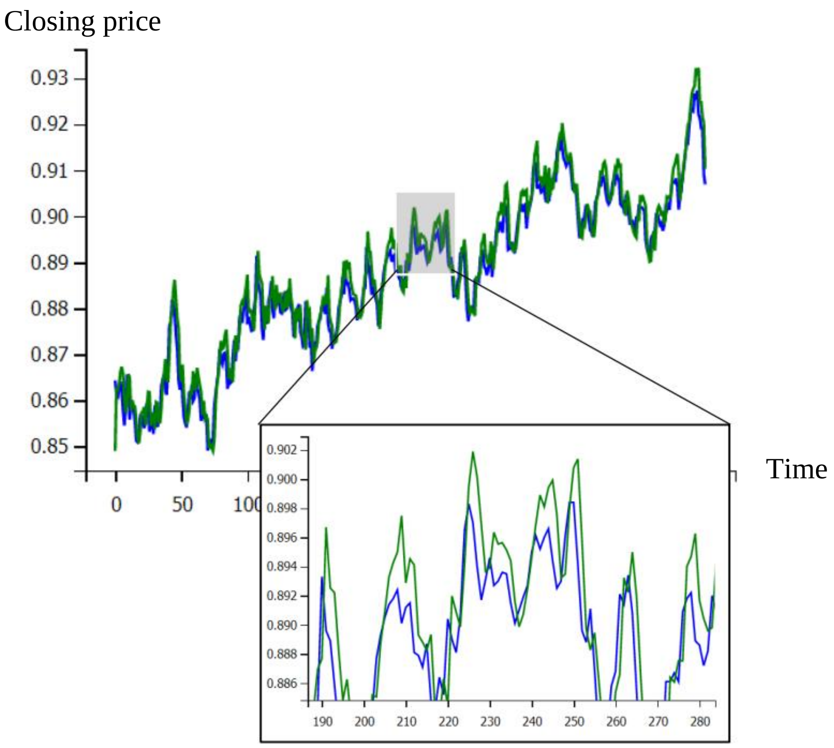

Figure 6. Persistence forecast problem

In general, Root Mean Square Error (RMSE) is chosen as an estimation technique almost for all Forex forecasting research. In the case of persistence forecast, RMSE can effectively assess the proximity between the prediction value and the actual value. Several studies, including [3], [20-30], produced persistence forecasting easily, particularly in financial time series forecasting. When the actual (blue line) and predicted (green line) time series are charted against one another, the predicted time series will appear to be one step behind the actual time series, as shown in Fig. 6. Although the persistence forecast outcome is close to the current price, it is not worth trading in the real world. The probability of receiving a correct direction signal is about 50%, which is the same as tossing a coin. This persistence forecast in complex models is caused by a number of factors. The selected time series problem, for example, is difficult to predict. Since fine-tuning parameters may be required, the chosen model cannot address problems with specific datasets.

4.5 EXPERIMENTAL RESULTS

Three experiments were carried out in this section to show that the proposed BERTFOREX-based FD and TI performed better other studies using either FD or TI (FD model and TI model). We also compare the effectiveness of cascading structures in TI and FD aggregation to the current TI and FD aggregation method based on heuristic aggregation rules.

No.	Research method	%correct direction signal	sensitivity	specificity	precision	NPV
1	FD model (FD+LSTM) [24]	49.21%	0.4612	0.5280	0.5144	0.4747
2	TI model (TI+LSTM) [25]	53.02%	0.5732	0.4859	0.5473	0.5123
3	FD and TI model based prediction [10]	51.12%	0.7440	0.2792	0.5097	0.52
4	**BERTFOREX**	**84.38%**	**0.8462**	**0.8457**	**0.8585**	**0.8324**
5	FD+BERT1+ANN	52.9%	0.5192	0.5425	0.5567	0.5049
6	TI+BERT1+ANN	61.61%	0.6538	0.5744	0.6296	0.6

Table 4. Experimental results

We constructed the last two experiments using FD+*BERT1*+ANN and TI+*BERT1*+ANN against FD based prediction method (FD model) and TI based prediction method (TI model), respectively, to show that latent relationship extraction of BERT could generate more accurate predictions than FD model and TI model. In order to make sure that all experiments were free of bias, the compared methods were used on the same dataset.

1. BERTFOREX vs. Either FD or TI based prediction method

The two experiments compared the performance of BERTFOREX and either FD or TI Forex forecasting models. When looking at the percentage of correct direction signals in Table 4, the TI model gets 53.02 percent, the FD model gets 49.21 percent, and the BERTFOREX model gets 84.38 percent.

The sensitivity, specificity, precision values, and NPV of three models are also assessed. With sensitivity, specificity, precision, and NPV of 0.8462, 0.8457, 0.8585, and 0.8324, respectively, BERTFOREX is the best. The TI model's values are 0.5732, 0.4859, 0.5473, and 0.5123, respectively, while the FD model's values are 0.4612, 0.5280, 0.5144, and 0.4747.

2. BERTFOREX vs. FD and TI aggregation based prediction method

The proposed cascade aggregation approach in BERTFOREX is compared to existing studies' heuristic aggregation rules to see how effective it is. BERTFOREX achieves higher measurement in all scores than FD and TI aggregation based prediction methods, as shown in experiments 3 - 4 in Table reftable4. BERTFOREX achieves a sensitivity value of 0.8462, while FD and TI aggregation-based prediction methods achieve a sensitivity value of 0.7440. As can be seen, BERTFOREX has the highest and most balanced sensitivity, specificity, precision, and NPV, making it suitable for real implementation.

3. FD+BERT1+ANN vs. FD based prediction method

We compare the results of FD+*BERT1*+ANN and FD-based prediction methods in this experiment. FD+*BERT1*+ANN achieves a percentage correct direction signals of 52.9 percent, sensitivity of 0.5192, specificity of 0.5425, precision of 0.5567, and NPV of 0.5049, while FD model achieves a percentage correct direction signals of 49.21 percent, sensitivity of 0.4612, specificity of 0.5280, precision of 0.5144, and NPV of 0.4747, as shown in Table reftable4. As can be seen, when compared to the FD model, FD+*BERT1*+ANN achieves a higher value in all measurements.

4. TI+BERT1+ANN vs. TI based prediction method

The purpose of this experiment was to see how well BERT performed in latent relation extraction within TI for generating prediction signals. The forecasting results of the TI+*BERT1*+ANN and TI-based prediction methods were compared. According to experiments 2 and 6 in Table reftable4, TI+*BERT1*+ANN produces a higher percentage of correct direction signals (61.61%) than the TI model (53.02%). TI+*BERT1*+ANN yields 0.6538, 0.5744, 0.6296, and 0.6 in terms of sensitivity, specificity, and NPV, respectively. In all measurement values, the TI model obtains only 0.5732, 0.4859, 0.5473, and 0.5123, which are lower than TI+*BERT1*+ANN.

CHAPTER 5 DISCUSSION

According to the experimental results in Table 4, BERTFOREX, which uses a cascading structure to aggregate TI and FD for Forex prediction, has good predictive performance as measured by the percentage of correct direction signals, sensitivity, and specificity. We'll go over the results of our experiments in more detail, as well as the reasons why our research produces the best prediction results.

5.1 FOREX PREDICTION BASED CHARACTERISTIC OF FD

We modify our work by picking up FD and *BERT1* to generate predictions compared to the prediction-based FD model [20] to prove that BERTFOREX handles the FD releasing problem better than other studies. FD+*BERT1*+ANN returns a higher percentage of correct direction, sensitivity, specificity, precision, and NPV than the FD model, as shown in Table 4. Despite the fact that Wijesinghe's study [20] was able to extract the relationship between FD and target price, the model's performance was affected by the FD releasing problem.

Because FD data, such as U.S. GDP, U.S. Housing Start Price, U.S. PPI, and U.S. Trade, are released quarterly/monthly, U.S. Nonfarm Payrolls are released on the first day of the month to predict a daily timeframe. As a result, nonfarm payrolls in the United States are missing values from day 2 to month end. All absent values were replaced with zero by Wijesinghe [20]. As a result, the inputs are missing values for the entire month. Their model is capable of capturing data relationships. The relationship quality, on the other hand, is poor because the input frequency is too low, and the prediction accuracy is poor.

The performance of BERTFOREX, on the other hand, is unaffected by this issue. We use BERT to transform FD absent values into FD latent representations, which represent the change in FD between quarters. Furthermore, the frequency of BERT output changes is the same as the target price. These two key characteristics enable the model to forecast currency prices more accurately than other models.

5.2 MODEL LEARNING

We compare the prediction result of *BERT1*+ANN in BERTFOREX to LSTM to show that BERTFOREX can capture and store time-series inner relationships better than other models. Two experiments were carried out to see how well these models predicted different types of input data.

First, we compared TI+*BERT1*+ANN to TI+LSTM for prediction-based TI. Then, as shown in Table reftable4, prediction-based FD was considered by comparing FD+*BERT1*+ANN with FD+LSTM. The experimental results show that TI+*BERT1*+ANN outperforms TI+LSTM in all metrics, including percentage of correct direction, sensitivity, and specificity, as shown in lines 2 and 6.

As shown in lines 1 and 5, FD+*BERT1*+ANN outperforms FD+LSTM in terms of percentage of correct direction, sensitivity, and specificity. Thus, regardless of the input data on TI or FD, *BERT1*+ANN outperforms LSTM. Furthermore, despite the low frequency of FD changing, BERT was able to capture the inner relationship and predict with a simple neural network better than the LSTM gate mechanism.

In general, the length of time series data grows with time, resulting in gradient vanishing and information loss in sequential models. With 4,455 data points, our Forex dataset is a non-stationary time series. The first-day data will be lost and unable to join the computation at the end of sequence forecasting due to the forget gate mechanism when using LSTM in time series forecasting. As a result, the model is unable to fully capture the characteristics of the data.

The proposed method, on the other hand, makes use of BERT's critical capabilities: bi-directional learning capability, self-attention mechanism, and positional embedding. Bidirectional learning enables the model to capture

Forex-related relationships in both the left-to-right and right-to-left directions. Some price patterns may be visible on both the left and right sides of the chart. For example, if a price rises sharply for a brief period of time, the price may fall due to the demand-supply law.

Date input sequences can interact with one another and determine the attention for each date using the self-attention mechanism. These date interactions and attention scores are aggregated in the self-attention outputs. Because the financial crisis occurred in 2003 and 2009, the USD/EUR price in November/1/2003 was most affected by October/28/2009. At this point, the attention score of date October/28/2009 provided by November/1/2003 was higher than the attention score of other dates; thus, the attention score was embedded on date position October/28/2009.

As a result, we used all attention scores as a day representation that would not be lost even if the price sequence was lengthened.

The structure of Forex, according to Forex data, is a time series structure where data order is critical for prediction. Because self-attention mechanisms alone cannot accurately reflect data sequence and position, we use a positional embedding process that allows BERTFOREX to capture the dependency between day elements at different positions in historical Forex data.

As a result, regardless of the input data type, BERTFOREX outperforms the existing sequential model as LSTM in Forex prediction.

5.3 UNDERLYING RELATIONSHIP EXTRACTION BETWEEN TI AND FD

To demonstrate that BERTFOREX's FD and TI relationship extraction is efficient, groundbreaking, and applicable in real life, we compare it to Yldrm et al. [5], which employs FD and TI based prediction rules. As shown in Table reftable4, BERTFOREX outperforms FD and TI-based prediction rules in terms of percentage correct direction, sensitivity, and specificity.

Between TI and FD, there is some relevant data that can be used to forecast price direction. For example, the United States' Gross Domestic Product (GDP) is a type of FD that reflects the country's production rates. Increased demand for US currency is a result of higher US GDP, which indicates that US products are in high demand. When demand for US currency is high, TI can detect a high volume of buying signals by increasing the RSI value. As a result, any model that captures the better TI-FD relationship will yield a more accurate prediction result.

While other studies combined TI and FD to create an accurate model, the current study [5] used a separate learning method to combine FD and TI before making a final prediction based on heuristic rules. As a result, the relationship between inner TI-TI and inner FD-FD is captured without taking into account other FD-TI relationships such as MA-EMA, MACD-RSI, and U.S.CPI-U.S.FFD. Note that the relationship between MA and the USCPI is not captured, despite the fact that there is a latent relationship between the two.

Furthermore, to make an accurate forecast, heuristic prediction rules can be used. Because the prediction rules are applied to the dataset 01/11/2017 - 01/02/2018, there may be some bias in the prediction. Although there were almost increasing trends during this time period, the prediction rules seemed to fit with the rising trend. Nonetheless, on the dataset 01/08/2017 - 01/04/2018, these rules may perform poorly, especially on the downtrend.

By adding FD latent representation to TI as additional weight, BERTFOREX can capture relationships across TI-FD. The latent relationship TI-FD will be created by learning FD and TI at the same time. BERTFOREX can quickly capture all changes between TI, FD, and the target closing price because this latent relationship has a frequency equal to the target price. Furthermore, the addition of FD to TI in BERTFOREX simulates real-world trader behavior, which involves making a decision, analyzing, and forecasting by comparing TI and FD side-by-side, resulting in more accurate predictions.

When the MA value is inserted into US GDP, *BERT2* is used to extract the latent relationship between MA and US GDP. BERTFOREX automatically interprets this as a significant increase in US GDP. As a result, the MA value gradually decreases, while the latent relationship between MA and US GDP becomes larger than in the previous time step.

All of the data obtained is more relevant to a price uptrend than a price downtrend, with BERTFOREX's output being a price-increasing signal.

As a result, BERTFOREX can generate a forecasting signal based on the real characteristics of TI and FD input data using TI-FD latent relationship extraction in *BERT2*. As a result, BERTFOREX achieves the highest value across all metrics, transforming it into a powerful Forex forecasting model.

CHAPTER 6 CONCLUSION

In this paper, we propose BERTFOREX, a Cascading model for Forex market forecasting that uses FD and TI and is based on BERT. By taking into account the actual characteristics of FD and TI, a new cascade aggregation procedure for FD and TI is introduced to enrich the Forex prediction performance, which is similar to real traders' behavior.

Because FD is released quarterly and has an impact on the price, we use BERT to learn latent relationships within FD and extract the FD, preventing BERTFOREX from having FD absent values. We also add the extracted FD to TI as a weighting factor to mimic the behavior of real-world traders, who primarily use TI rather than FD to generate forecasts. The TI-FD relationship is extracted using the consecutive BERT and then used as input for the prediction task.

For exchange rate forecasting, BERTFOREX aggregated representation provides more accurate results than other Forex forecasting techniques. Because the BERTFOREX output representation is almost entirely made up of FD and TI latent patterns from the past and is generated from the better extracted hidden relationship of FD, it is made up of FD and TI characteristics. These claims are backed up by our experimental findings, which show that our BERTFOREX model outperformed many other Forex forecasting models.

In future work, we plan to continue working on the high-frequency Forex prediction model. Hourly forecasts, for example. External factors that change the price quickly must be handled by the model. When dealing with noisy and chaotic time series, high-frequency forecasting is a particularly difficult task.

REFERENCES

1. Qi, M. and Y. Wu, *Nonlinear prediction of exchange rates with monetary fundamentals.* Journal of Empirical Finance, 2003. **10**(5): p. 623-640.
2. Eng, M.H., et al. *Forecast forex with ANN using fundamental data*. in *2008 International Conference on Information Management, Innovation Management and Industrial Engineering*. 2008. IEEE.
3. Yao, J. and C.L. Tan, *A case study on using neural networks to perform technical forecasting of forex.* Neurocomputing, 2000. **34**(1-4): p. 79-98.
4. AbuHamad, M., M. Mohd, and J. Salim, *Event-driven business intelligence approach for real-time integration of technical and fundamental analysis in forex market.* Journal of Computer Science, 2013. **9**(4): p. 488.
5. Yıldırım, D.C., I.H. Toroslu, and U. Fiore, *Forecasting directional movement of Forex data using LSTM with technical and macroeconomic indicators.* Financial Innovation, 2021. **7**(1): p. 1-36.
6. Emam, A. *Optimal artificial neural network topology for foreign exchange forecasting*. in *Proceedings of the 46th Annual Southeast Regional Conference on XX*. 2008.
7. Handa, R., A. Shrivas, and H. Hota, *Financial time series forecasting using back propagation neural network and deep learning architecture.* Int. J. Recent Technol. Eng, 2019. **8**: p. 3487-3492.
8. Ifadhila, I., I. Haanurat, and A. Jaya, *Fundamental indicators analysis toward foreign exchange (forex) prediction using neural network method-radial basis function (NN-RBF).* Prosiding CELSciTech, 2018. **3**.
9. Johansson, S. and S. Nafar, *Effective Sampling and Windowingfor an Artificial Neural Network Model Used in Currency Exchange Rate Forecasting*. 2017.
10. Ling, J., A.K. Tsui, and Z. Zhang. *Forecast foreign exchange with both linear and non-linear models coupled with trading rules for selected currencies*. in *The 21st International Congress on Modelling and Simulation (MODSIM2015), Gold Coast Convention and Exhibition Centre, Broadbeach, Queensland, Australia*. 2015.
11. Shiao, Y.C., et al. *Modeling and Prediction of Time-Series-A Case Study with Forex Data*. in *2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST)*. 2019. IEEE.
12. Tenti, P., *Forecasting foreign exchange rates using recurrent neural networks.* Applied Artificial Intelligence, 1996. **10**(6): p. 567-582.
13. Maknickienė, N. and A. Maknickas. *Application of neural network for forecasting of exchange rates and forex trading*. in *The 7th international scientific conference" Business and Management*. 2012.
14. Hiew, J.Z.G., et al., *BERT-based financial sentiment index and LSTM-based stock return predictability.* arXiv preprint arXiv:1906.09024, 2019.
15. Kumar, H.P. and B.S. Patil. *Forecasting volatility trend of INR USD currency pair with deep learning LSTM techniques*. in *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*. 2018. IEEE.

16. Yong, Y.L., D.C. Ngo, and Y. Lee. *Technical indicators for forex forecasting: a preliminary study*. in *International Conference in Swarm Intelligence*. 2015. Springer.
17. Vaswani, A., et al. *Attention is all you need*. in *Advances in neural information processing systems*. 2017.
18. Devlin, J., et al., *Bert: Pre-training of deep bidirectional transformers for language understanding.* arXiv preprint arXiv:1810.04805, 2018.
19. Huisu, J., et al. *Predicting bitcoin prices by using rolling window LSTM model*. in *Proceedings of the KDD data science in Fintech Workshop, London, UK*. 2018.
20. WIJESINGHE, S., *Time Series Forecasting: Analysis of LSTM Neural Networks to Predict Exchange Rates of Currencies.*
21. Qu, Y. and X. Zhao. *Application of LSTM neural network in forecasting foreign exchange price*. in *Journal of Physics: Conference Series*. 2019. IOP Publishing.
22. Kondratenko, V. and Y.A. Kuperin, *Using recurrent neural networks to forecasting of forex.* arXiv preprint cond-mat/0304469, 2003.
23. Kamruzzaman, J. and R.A. Sarker. *Forecasting of currency exchange rates using ANN: A case study*. in *Proceedings of the 2003 international conference on neural networks and signal processing*. 2003. IEEE.
24. Dash, R., *Performance analysis of an evolutionary recurrent Legendre Polynomial Neural Network in application to FOREX prediction.* Journal of King Saud University-Computer and Information Sciences, 2020. **32**(9): p. 1000-1011.
25. Dash, R., *An improved shuffled frog leaping algorithm based evolutionary framework for currency exchange rate prediction.* Physica A: Statistical Mechanics and its Applications, 2017. **486**: p. 782-796.
26. Dobrovolny, M., et al., *Forecasting of FOREX Price Trend using Recurrent Neural Network-Long short-term memory.* 2020.
27. Papatsimpas, M.G., I. Lykogiorgos, and K.E. Parsopoulos. *FOREX Trading Model Based on Forecast Aggregation and Metaheuristic Optimization*. in *11th Hellenic Conference on Artificial Intelligence*. 2020.
28. Cao, W., et al., *A deep coupled LSTM approach for USD/CNY exchange rate forecasting.* IEEE Intelligent Systems, 2020. **35**(2): p. 43-53.
29. Zhelev, S. and D.R. Avresky. *Using LSTM neural network for time series predictions in financial markets*. in *2019 IEEE 18th international symposium on network computing and applications (NCA)*. 2019. IEEE.
30. Jia, H., *Deep learning algorithm-based financial prediction models.* Complexity, 2021. **2021**.

VITA

NAME	Arisara Pornwattanavichai
DATE OF BIRTH	21 December 1995
PLACE OF BIRTH	Bangkok, Thailand
INSTITUTIONS ATTENDED	Bachelor of Science in Computer Science
HOME ADDRESS	449 Maitrijit road Pomprab Pomprab Bangkok 10100
PUBLICATION	- Enhanced Tweet Hybrid Recommender System Using Unsupervised Topic Modeling and Matrix Factorization-Based Neural Network - Tweet Hybrid Recommendation Based on Latent Dirichlet Allocation
AWARD RECEIVED	- Best Paper Awards 2018