Machine Learning-Based Stock Market Price Prediction

pavan reddy narra
Prpavan80@gmail.com

Abstract

Every day large sums of money flow in and out of the trading market. Investing in the stock market made some success and for others it's a disaster. Stock market price prediction is a challenging and complex task due to its dynamic environment and the price of a stock depends on lots of factors. predicting the movement of the stock by using Machine learning may help the investors to take correct decisions in the right direction and mitigate the losses

Table Of Contents

List OF Tables Used

List Of Figures Used

CHAPTER1: INTRODUCTION

1.1. Background

The stock market is depending on multiple political, social and economic factors.so making predictions about the stock price is difficult. There are different kinds of investors like day traders, momentum trading, short term trading, long term trading. According to the paper (Fernando Chague 2019),97% of people who do day trading lose money. Most of the retail investors lose money in the market due to many reasons (Jitender Kumar 2022) like lack of experience, limited knowledge of stock and economics, overconfidence and bias. According to paper (Heliyon 2021) most of traders are not using stop losses and the tendency to hold on to losing trade for extended periods and quickly exit profitable positions to secure gains is typical behavior shown by novice traders. So, by developing the Machine learning models by using technical and fundamental analysis might help them to understand the momentum of the market to execute the proper decision.

1.2. Problem Statement OR Related Research OR Related Work

Predicting the stock price is one if the complex task because of its nonlinear relationship and involvement of lot of factors. There are many research papers who attempted to solve the problem by using time series forecasting algorithms and deep learning algorithms. Few papers only considered the technical analysis while few papers combined both technical and fundamental analysis. The work by (Matin N. Ashtiani 2023) used fundamental and technical analysis by using different models LSTM, CNN and SVM to predict the price and find the strong correlation with sentimental analysis and stock price fluctuations. The paper written by (Sridhar and Sanagavarapu 2021)used multi head attention to predict the price of dodge coin by using historical price and attained the accuracy higher when compared to other state of art algorithms. The papers (Yijiao Liu 2022) & (Sabu 2022) proposed using frequency decomposition for noise reduction by using CEEMD, Encoders and then feeding into LSTM/GRU and had better accuracy than compared to plain LSTM .The papers (Chengyu Li n.d.)& (Ming-Che Lee 2022) used attention mechanism but only considered the technical data for prediction. The paper (M.

Reza Pahlawan 2021)uses the macroeconomics to predict the price by using RNN. By referring to the past works want to use the recent transformer architecture block recurrent transformer which is a combination of LSTM and Transformers to predict the dynamic nature of the price movement as they are more suitable for time series analysis. With reference to previous research, I plan to utilize the Block-Recurrent Transformer (DeLesley Hutchins 2022), a combination of LSTM and Transformers by giving the by including macroeconomic factors, fundamental and technical data to predict the dynamic nature of price movements in time series analysis and to check if it can able to achieve better accuracy when compared to earlier models as proposed in the research papers. This architecture excels at capturing temporal dependencies and sequential patterns, making it suitable for time series analysis.

1.3. Research Questions (If any)

What are the relevant data sources and features in order to extract the dynamic nature of the market?

What are state of art algorithms to trace out the nonlinear relationship?

What kind of evaluation metrics to use?

1.4. Aim and Objectives

The main goal of this research is to propose a model to predict the price of the stock by considering technical and fundamental data of the stock which includes historic price data and sentimental analysis by using attention mechanism.

The following are the objectives of this project study:

- To understand the correlation between the sentimental analysis and price changes
- To understand the patterns from historic data
- To compare the accuracy with other state of art algorithms
- To evaluate the performance of the model

1.5. Significance of the Study

The study focuses on developing a model to predict the momentum of stock prices of the stocks on which it is trained. The research may be helpful for investors and beginners who are entering the market. It gives an idea of the market and helps them to either enter or exit the stock by minimizing the losses.

1.6. Scope of the Study

The study focuses on a few USA stocks. Only a few stocks were considered because collecting the data is a tedious task and news articles for those stocks are for sentimental analysis.

1.7. Research Plan

Table 1: Project Plan

Topic	Start Date	End Date	Days
Literature Review	2023-04-01	2023-05-01	30
Dataset source selection	2023-05-01	2023-05-15	14
Thorough Literature review to Understand different model architectures and functionality	2023-05-15	2023-05-28	13
Looking for different data sources due to change in functionality of twitter as web scraping become difficult to change in the UI of Twitter	2023-05-28	2023-06-05	14

Going through the Literature review again because of using news articles instead of social media data, as Twitter changed UI and stock twits stop giving new API developer keys	2023-06-05	2023-06-17	12
Interim Report	2023-06-17	2023-07-03	7
Data Collection and model development	2023-07-03	2023-07-15	12
Results and analysis ,any modifications on Interim report ,changes in model either inputs or hyperparameters tunning	2023-07-15	2023-07-30	25
Further improvements if needed, Dissertation report preparation and presentation slides preparation	2023-08-01	2023-08-24	24

Chapter2: Literature Review

2.1. Introduction

This review provides a comprehensive overview of the research related to stock price prediction using machine learning techniques. The literature review will discuss the various machine learning techniques used to predict stock prices, the various datasets used, and the results of the studies.

First, a variety of machine learning techniques such as recurrent neural networks, deep learning, frequency decomposition, self-attention transformers, and hybrid information mixing modules are discussed. These techniques have been used to create models that predict stock prices with varying levels of accuracy. Alejandro Lopez-Lira and Yuehua Tang (2023) studied the use of the ChatGPT model to predict stock price movements, while Burra et al. (2023) used zero-shot sentiment classification, and Chengyu Li and Guoqi Qian (n.d.) used a frequency decomposition-based GRU transformer neural network. Other studies have used hybrid information mixing modules (Choi et al., 2023) and block-recurrent transformers (DeLesley Hutchins et al., 2022).

Second, various datasets have been used in these studies. These datasets include both historical stock market data and news articles, which can be used to capture sentiment regarding stocks and their movements. Fernando Chague et al. (2019) used data from the Brazilian stock market, while Heliyon (2021) used data from the Indian stock market. Hu (2021) used data from the Shanghai Stock Exchange, and Jitender Kumar and Neha Prince (2022) used data from the National Stock Exchange of India.

This literature review has provided an overview of the research related to stock price prediction using machine learning techniques. The review discussed the various machine learning techniques used to create models for stock price prediction, the datasets used in the studies, and the results of the studies. In the below different models mentioned in the reference papers are studied, it first focuses on recurrent networks and then hybrid models using transformers, attention and recurrent networks.

2.2. Stock Price Prediction Using RNN, LSTM and GRU

2.2.1. Stock market price prediction by using RNN

The paper “Stock Price Forecast of Macro-Economic Factor Using Recurrent Neural Network” by M. Reza Pahlawan, Edwin Riksakomara, and Raras Tyasnurita (2021) has provided a comprehensive overview of the utility of recurrent neural networks (RNN) for stock price forecasting. The authors provide a detailed discussion of the various methods used to build stock price forecasting models, such as time-series analysis, econometric analysis, and machine learning techniques. The authors also provide an extensive review of the literature on stock price forecasting, including a focus on the use of RNNs.

The paper highlights the unique advantages of using RNNs for stock price forecasting. Specifically, the authors discuss the ability of RNNs to capture complex patterns and trends in the data, as well as their advantages over traditional methods such as econometric analysis. Additionally, the paper describes the various techniques for constructing RNN models, such as the use of long short-term memory (LSTM) and gated recurrent units (GRUs). The authors also discuss the various datasets that are available for stock price forecasting, such as the Kaggle dataset.

The paper identifies several research gaps in the field of stock price forecasting. First, the authors note that there is a lack of studies that use RNNs to forecast stock prices over long-term horizons. Second, the authors suggest that more research is needed to explore the potential of using alternative machine learning techniques, such as deep learning, for stock price forecasting. Third, the authors note that the current literature on stock price forecasting does not provide sufficient guidance on the best practices for constructing RNN models.

The paper also identifies several potential challenges associated with the use of RNNs for stock price forecasting. First, the authors suggest that RNNs are not well-suited for making short-term predictions due to their reliance on past data. Second, the authors note that the ability of RNNs to capture complex patterns in the data is limited, as they are unable to capture non-linear relationships. Lastly, the authors suggest that the datasets used for stock price forecasting are often limited in both size and quality.

The dataset used in the paper includes macroeconomic factors such as the Consumer Price Index (CPI), the Unemployment Rate, the Money Market Rate (MMR), and the Stock Price Index (SPI). The dataset u is taken from the Indonesian Stock Exchange (IDX). It consists of daily closing prices of 15 stocks from the mining and energy sectors. The evaluation metrics used in the paper are mean absolute error (MAE) and root mean square error (RMSE). These metrics are used to measure the accuracy of the forecasts generated by the recurrent neural network model. The results are then compared to the ARIMA and Prophet models to determine the most accurate model.

In conclusion, the paper by M. Reza Pahlawan, Edwin Riksakomara, and Raras Tyasnurita (2021) provides an extensive review of the literature on stock price forecasting. The authors identify several research gaps in the field, including the lack of studies that use RNNs to forecast stock prices over long-term horizons, the need for more research on the potential of using alternative machine learning techniques for stock price forecasting, and the lack of guidance on best practices for constructing RNN models. Additionally, the authors discuss several potential challenges associated with the use of RNNs for stock price forecasting, such as their reliance on past data and their inability to capture non-linear relationships.

2.2.2. LSTM and GRU model for stock market prediction

The paper “An LSTM and GRU based trading strategy adapted to the Moroccan market.” by Touzani, Y. & Douzi (2021) proposes stock price forecasting with LSTM and GRU architectures. The focuses on the different methods used, comparison of different methods, evaluation metrics, scope, and research gaps.

Stock price predictions have been made using a variety of machine learning and deep learning algorithms. Due to their adaptability for sequential input, Long-Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) designs are two of the most popular recurrent neural network (RNN) types.In the study, LSTM and GRU models are used to successfully forecast stock prices in a number of markets, including the US, Brazil, South Africa, Iran, India, Russia, Mexico, the UK, Hong Kong, Indonesia, and Ghana.

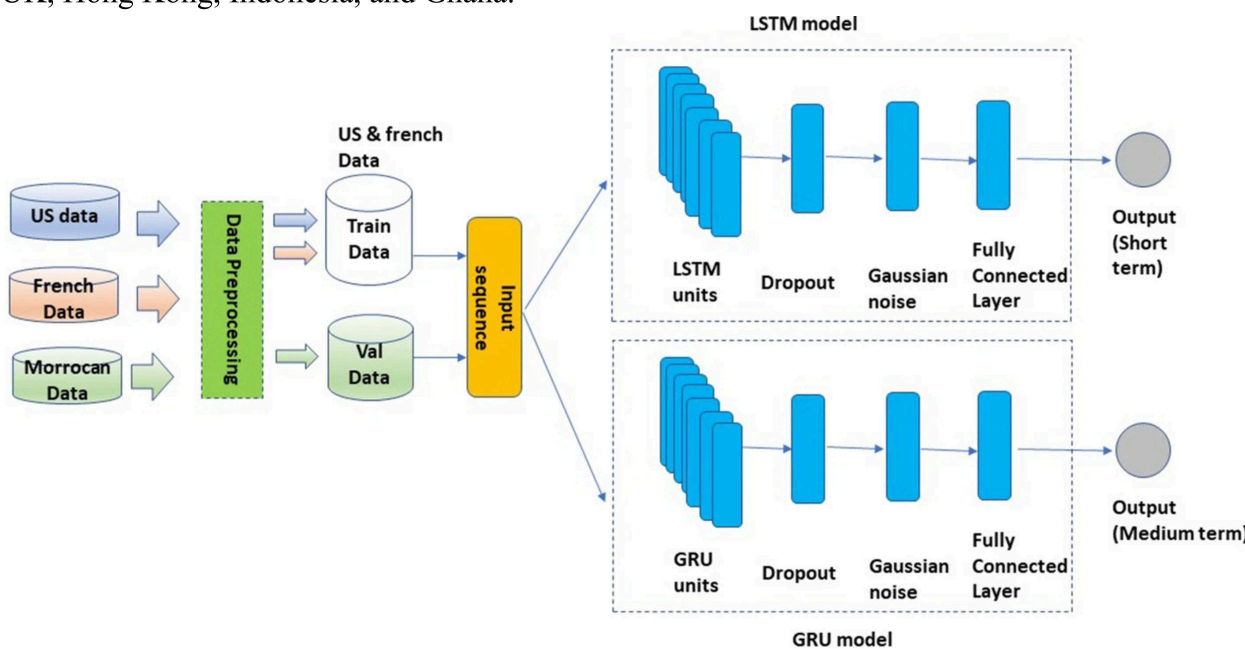


Fig 2.1. Lstm and Gru model for stock price prediction

The evaluation metrics used in the different papers vary, but most commonly include mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE). Different methods, such as empirical wavelet transform (EWT) and empirical mode decomposition (EMD) have been used to decompose closing price data. Other methods, such as particle swarm optimization (PSO) and outlier robust extreme learning machine (ORELM) have been used to fine-tune model hyperparameters and correct errors.

The suggested method produces results that are quite encouraging and outperforms the performance of indexes that are utilised as benchmarks in the local market. The annualised return of our recommended strategy during the test period was 27.13%, as opposed to the Moroccan All

Share Indice's (MASI) annualised return of 0.43%, the annualised return of the distribution sector at 15.24%, and the annualised return of the pharmaceutical sector at 19.94%.

The scope of the papers discussed in this literature review focused mainly on short-term and medium-term stock price forecasting. Furthermore, the research gaps discussed in this literature review include a lack of studies devoted to the Moroccan stock market, and the need for a trading strategy designed for the Moroccan market that can outperform the performance of indices used as benchmarks in the local market.

2.3. Using Deep auto Encoders and Decoders

Need for auto encoders- An autoencoder is a type of artificial neural network used to learn efficient data coding in an unsupervised manner. In the paper "A cooperative deep learning model for stock market prediction using deep autoencoder and sentiment analysis" by Sabu, KS Rekha.MK. 2022, the authors explain that the use of autoencoders can help to extract the essential features from the large amount of data present in the stock market and improve the accuracy of the stock prediction. The deep autoencoder can help to reduce the dimensionality of the data and extract only the important features for the prediction. Moreover, the autoencoder can help to capture nonlinear features in the data which may be difficult to represent using traditional models. Thus, the use of autoencoders can help to improve the performance of stock market prediction models.

Methodology- An autoencoder receives the historical stock market data, which could include noise. The autoencoder is a self- or unsupervised learning model that can encode and decode data without the aid of a human or the labeling of the data. The encoder converts the input data into a representation with fewer dimensions. With as little information loss as possible, the decoder attempts to recreate the original data from the lower-dimensional representation.

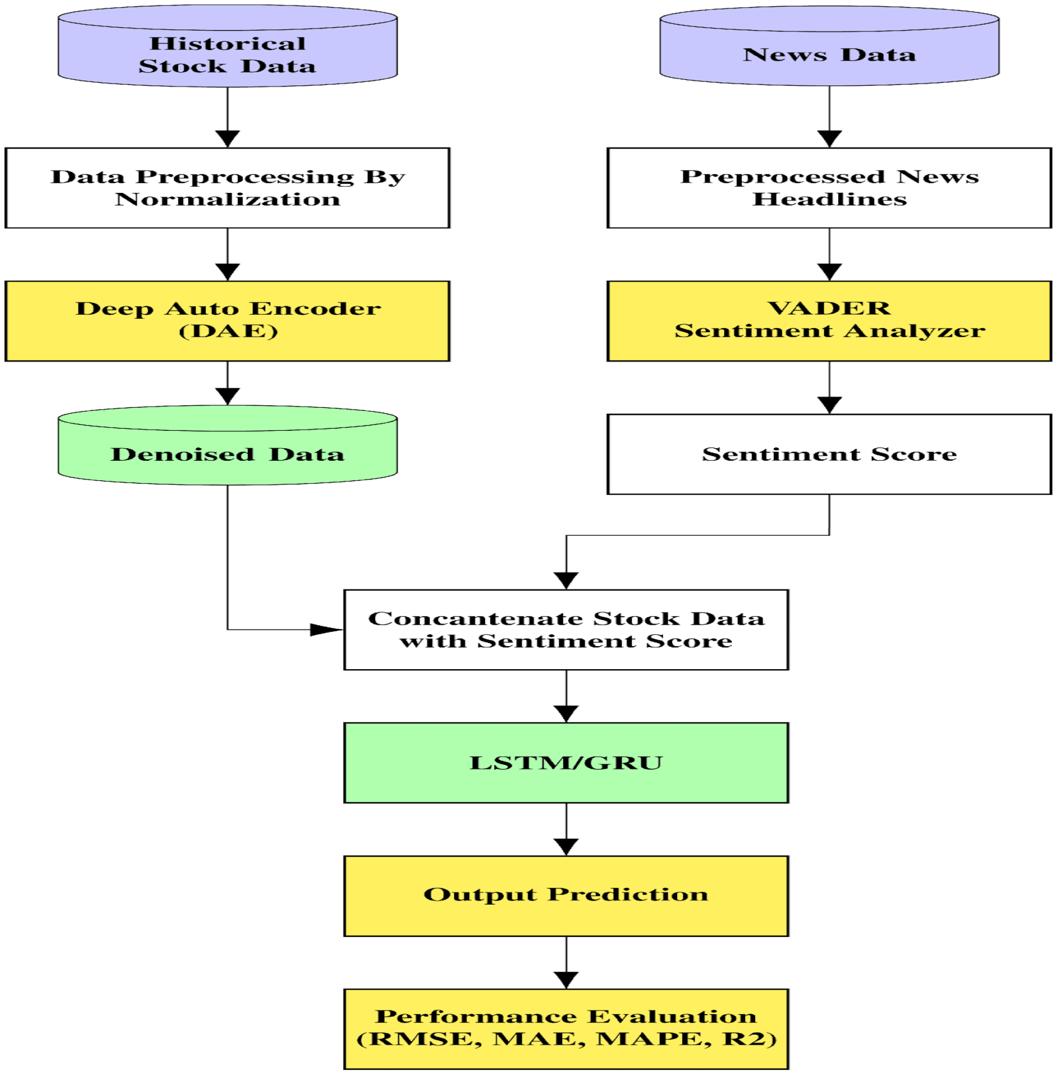


Fig 2.2 Lstm/Gru with autoencoder and sentimental analysis

To denoise the noisy historical stock market data, the data is passed through a DAE. A decoding half and an encoding half of two symmetrical deep belief networks make up the DAE. The goal is to recover the original data from a lower-dimensional representation with as little information loss as possible.

The LSTM or GRU model is then fed the autoencoder's denoised data. In parallel, sentiment analysis of company-related news for the relevant time period is carried out. The LSTM/GRU model incorporates the sentiment data, adding additional input for stock prediction.

Metrics like RMSE, MAE, MAPE, and R2 are used to assess the performance of the models. These metrics assess the models' precision and propensity for prediction. According to the study, incorporating news sentiment significantly enhances stock price prediction accuracy. In a collaborative setting, the LSTM model outperforms the GRU model with news sentiments.

The outcomes demonstrate superior performance over the benchmark models for the proposed DAE-LSTMSA and DAE-GRUSA models, as evidenced by lower error rates and higher R2 values. The accuracy of stock market forecasting is increased by the inclusion of news headlines and sentiment analysis. According to the study, incorporating sentiment analysis into deep learning models improves their performance and has a significant impact on stock market prediction. The proposed models outperform the benchmark LSTM and GRU models in terms of RMSE, MAE, MAPE, and R2 metrics, as shown by the comparison with cutting-edge models for various stock datasets (HMC, ORCL, and INTU).

2.4. Deep Learning based Stock Market Prediction considering Covid-19 as a Feature

The proposed study by Mondal, Bhaskar, and Asimkiran Dandapat (2022) aims to develop a machine learning and deep learning-based stock market prediction model that considers the effect of Covid-19. This will build on previous studies that have explored the potential of applying machine learning and deep learning to stock market prediction and will also provide a more comprehensive approach that takes into account the effects of Covid-19 on stock market performance.

Methodology

The model uses different algorithms and compares accuracy. Decision Tree, Gradient boosting, Lasso and Ride Regression, Random Forest, SVR, LSTM, GRU. Understanding the effects of the COVID-19 pandemic on the stock market is one of the study's main goals. Researchers want to know how the market changed during the pandemic and if it will return to normal in the future. The market's behaviour during a crisis and its prospects for recovery are revealed by this visualisation.

The paper's second contribution is to choose the optimum model for forecasting the financial market based on the considered inputs. In order to discover which ML and DL model is most effective in forecasting the stock market, researchers analyse and assess a variety of ML and DL models. The study's goal is to offer a useful tool for those who need it by determining the most precise and efficient model.

Datasets

The dataset has 21 features and includes the following.

Current Day NIFTY 50 Return and Miscellaneous: The NIFTY50 index's daily returns as well as other unspecified factors are included in this category. The COVID-19 dataset from the Centre for Systems Science and Engineering (CSSE) at JHU and NSE India serve as the data sources for these aspects. Data from COVID-19 span the months of January 2020 and February 2021.

Economic and Financial Features of the Raw Data (Macroeconomics: Crude, Gold, and Foreign Exchange Value): The macroeconomic factors that affect the price of crude oil, gold, and foreign exchange rates fall under this category. In contrast to the foreign exchange data, which is sourced from finance.yahoo.com, the data for gold and crude oil is gathered from mcxindia.com.

Economic and Financial Features of the Raw Data (Major Indices): Major indices including the Hang Seng Index, Shanghai Stock Exchange, Standard and Poor 500, NIKKEI 225, Strait Times, and FTSE 100 fall under this group. These indices' data sources include finance.yahoo.com and Bloomberg.com.

Results and analysis

The paper uses the Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) metrics to evaluate the performance of its models.
The results of the paper showed that the model which utilized both the machine learning and deep learning approaches was the most effective at predicting stock prices with a MAPE score of only 4.9%.
The paper also used a sensitivity analysis to evaluate how the model performed under different scenarios. This analysis showed that the model was able to accurately predict stock prices under varying conditions, demonstrating its effectiveness.

News-based intelligent prediction of financial markets using text mining and machine learning:

A literature review of the paper "News-based Intelligent Prediction of Financial Markets Using Text Mining and Machine Learning: A Systematic Literature Review" by Matin N. Ashtiani and Bijan Raahemi (2023) provides a comprehensive overview of the methods, evaluation metrics, scope, and research gaps associated with the application of text mining and machine learning techniques to financial markets.

The authors discuss the use of text mining and machine learning for various tasks related to financial markets, such as predicting stock prices, detecting stock market anomalies, extracting financial sentiment from news, and financial event detection. They also compare various methods and algorithms used for these tasks, including traditional methods such as linear regression, support vector machines, and decision trees, as well as more recent machine learning algorithms such as recurrent neural networks, long short-term memory networks, and deep learning.

The authors also discuss the evaluation metrics used to assess the performance of these methods. These metrics include accuracy, precision, recall, F1 score. Additionally, the authors provide an overview of the scope and limitations of the current research on news-based intelligent prediction of financial markets. They highlight the need for further research on the integration of natural language processing and machine learning techniques, as well as the use of transfer learning and reinforcement learning for financial markets.

Finally, the authors discuss some potential research gaps in the field of news-based intelligent prediction of financial markets. These include the need for better algorithms and evaluation metrics to better capture the non-linear relationships between news and financial markets, as well as the need for more comprehensive datasets.

In conclusion, this systematic literature review provides a comprehensive overview of the methods, evaluation metrics, scope, and research gaps associated with the application of text mining and machine learning to financial markets.

2.5. Stock Price Prediction Using Attention Mechanisms and Transformers

2.5.1. Using multi-head self-attention transformer

The paper by Sridhar, Shashank, and Sowmya Sanagavarapu (2021) examines the use of multi-head self-attention transformer (MAAT) for predicting the price of Dogecoin. The authors recognize that Dogecoin, a cryptocurrency, is rapidly becoming a more popular form of digital currency, and that its price can be highly volatile. As such, they seek to develop a predictive model that can accurately perform Dogecoin price prediction.

The authors first discuss the existing literature on cryptocurrency price prediction, noting that while various machine learning models have been proposed, the accuracy of these models is still limited. They then present their proposed MAAT model, describing its architecture and its ability to capture temporal dependencies in the data. The authors claim that their model outperforms traditional machine learning models in terms of accuracy and performance.

The methodology used in this research is a multi-head self-attention transformer for Dogecoin price prediction. The transformer architecture is based on the transformer neural network model, which is an encoder-decoder architecture that uses attention mechanisms to improve the performance of the network. The transformer architecture is composed of multiple layers of self-attention, which are used to capture long-term dependencies in the data. The transformer uses multiple heads of self-attention, which allow the model to focus on different parts of the data at the same time. The model is then trained using a supervised learning approach, where the model is trained on a dataset of Dogecoin prices and then used to predict future prices.

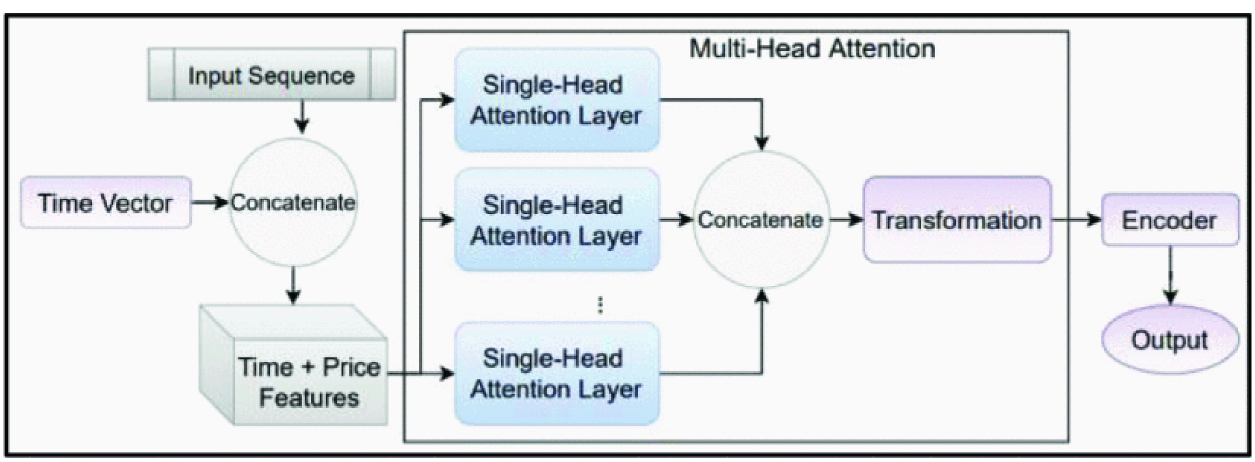


Fig 2.3. Multiheaded Attention for dogecoin price prediction

In order to evaluate their model, the authors conducted experiments using a dataset containing Dogecoin prices from January 2017 to August 2020. They compared the results of their MAAT model to those of traditional machine learning models, such as Random Forest and Support Vector Machines, as well as other deep learning models, such as LSTM and GRU. The results showed that the MAAT model was able to outperform all the other models in terms of accuracy and provided more accurate predictions of Dogecoin prices.

2.5.2 stock price prediction using Attention based BILSTM

This paper by Ming-Che Lee, Jia-Wei Chang, Sheng-Cheng Yeh, Tsorng-Lin Chia, Jie-Shan Liao, and Xu-Ming Chen(2022) presents an approach to stock trading strategies that relies on attention-based BiLSTM and technical indicators. The authors used a dataset of Taiwanese stocks, along with several technical indicators such as moving average convergence/divergence (MACD), relative strength index (RSI), and stochastic oscillator (SO) to create their trading strategies. The authors then conducted a performance analysis of their strategies using various metrics, such as Sharpe ratio, maximum drawdown, and annual return.

The authors find that their strategies outperform the traditional buy-and-hold strategy in terms of both risk and return. The authors also find that their strategies can capture the market trend more accurately than the traditional buy-and-hold strategy. Furthermore, the authors find that their strategies can handle market volatility better than traditional strategies.

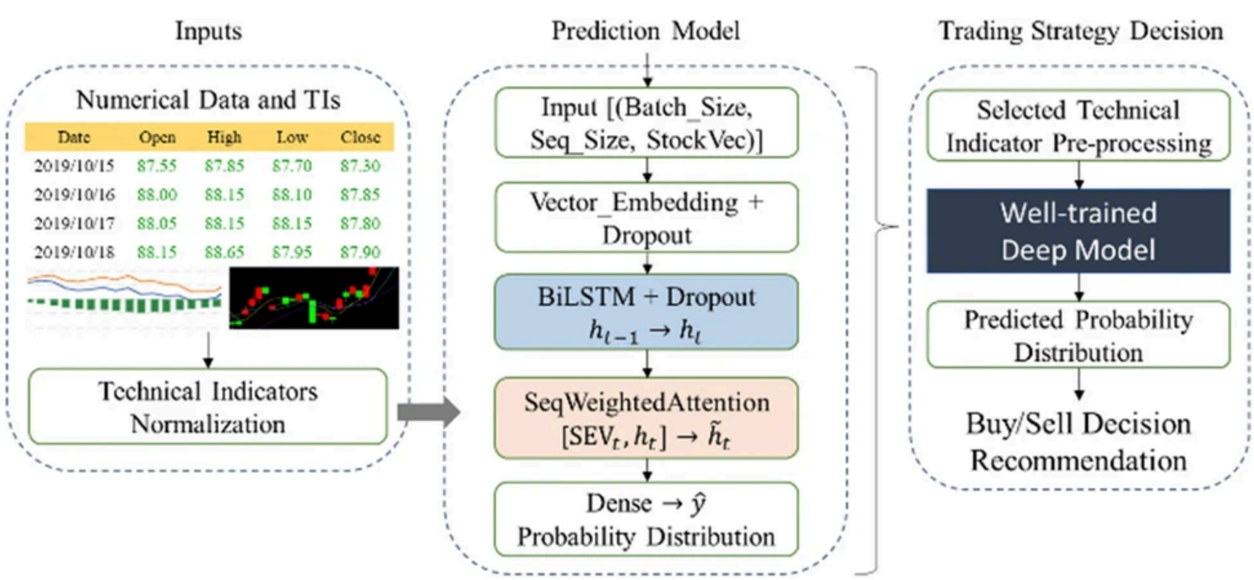


Fig 2.4. Attention Based Bilstm

In the model the sequence input which is concatenated stock vector is given as an input to the Bilstm model. The stock vector contains the trading prices and technical indicators. The bilstm output the weighted hidden stated vectors which are calculated on the input sequence within the rolling window. These hidden states are given as a input to the attention mechanism which generates the context vectors and it can be here helpful to calculate the temporal relationship between the rolling windows. The context vector is given to the dense layer to predict probability which gives either buy or sell.

The authors' results suggest that their attention-based BiLSTM and technical indicators-based trading strategies are a promising approach for stock trading. The use of these strategies could provide investors with better returns and lower risk than traditional strategies. Furthermore, these strategies could be used to capture market trends more accurately and handle market volatility better.

The authors also suggest several future studies that could be conducted to further explore the use of attention based BiLSTM and technical indicators in stock trading strategies. For example, further research could be conducted to investigate the impact of different technical indicators on trading strategies. Additionally, the authors suggest that more comprehensive datasets could be used to increase the accuracy of the strategies.

2.5.3. By Using a Hybrid Model Using CEEMD, Entropy, GRU, and History Attention for Intraday Stock Market Forecasting.

The paper by Yijiao Liu, Xinghua Liu, Yuxin Zhang,Shuping Li(2022) titled CEGH: A Hybrid Model Using CEEMD, Entropy, GRU, and History Attention for Intraday Stock Market Forecasting is an investigation into the use of a hybrid model for stock market forecasting. This paper proposes a novel model, CEGH, which combines CEEMD, Entropy, GRU, and History Attention. The authors argue that this model offers a more accurate prediction for intraday stock prices than existing models, as it considers multiple factors such as market trends, sentiment, and past prices.

The authors begin by introducing the concept of intraday stock price forecasting and its importance to investors. They then describe the existing models used for intraday stock price forecasting and their limitations. Next, the authors discuss the components of the proposed model, CEGH. The authors explain that the CEGH model is composed of four components: CEEMD, Entropy, GRU, and History Attention. CEEMD decomposes the input time series into several intrinsic mode functions which capture the oscillatory behavior of the underlying series. Entropy is used to measure the degree of disorder of the data. GRU is used to capture the internal temporal dependencies within the data, while History Attention captures the long-term trends in the data.

The inputs are given to the Ceemd and it decomposes the given input into different frequency decompositions. For example, imf1 will capture high frequency signals like highest price for the given rolling window period, imf2 captures closing prices and imf3 captures monthly trends.

These are then fed to entropy denoising which removes the noise from the signal. It separates the high frequency noise from the lower frequency signal. The lower frequency signal is here grouped into four groups and different models were used for each group. Here Gru and history attention was used to capture the temporal relations within the rolling window and between windows. The Gru outputs the hidden states and then they are given to the attention to capture long term dependencies. The output from the models is ensembled and gives the output probability.

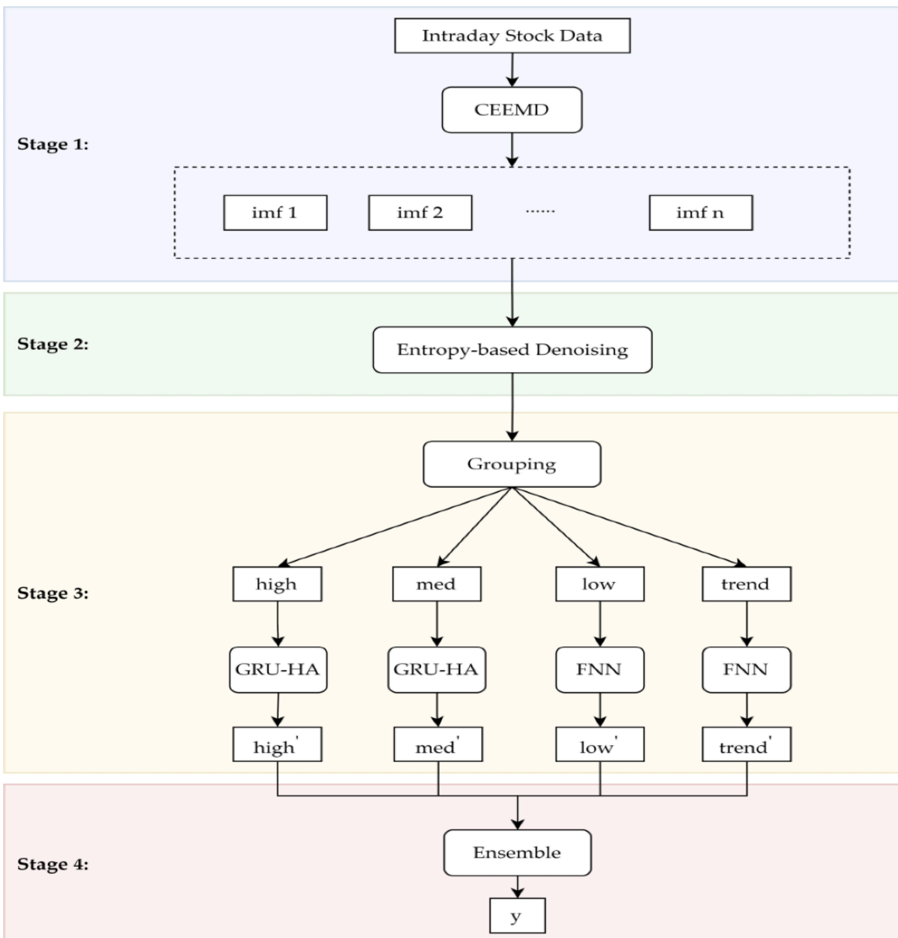


Fig2.5. CEGH Architecture

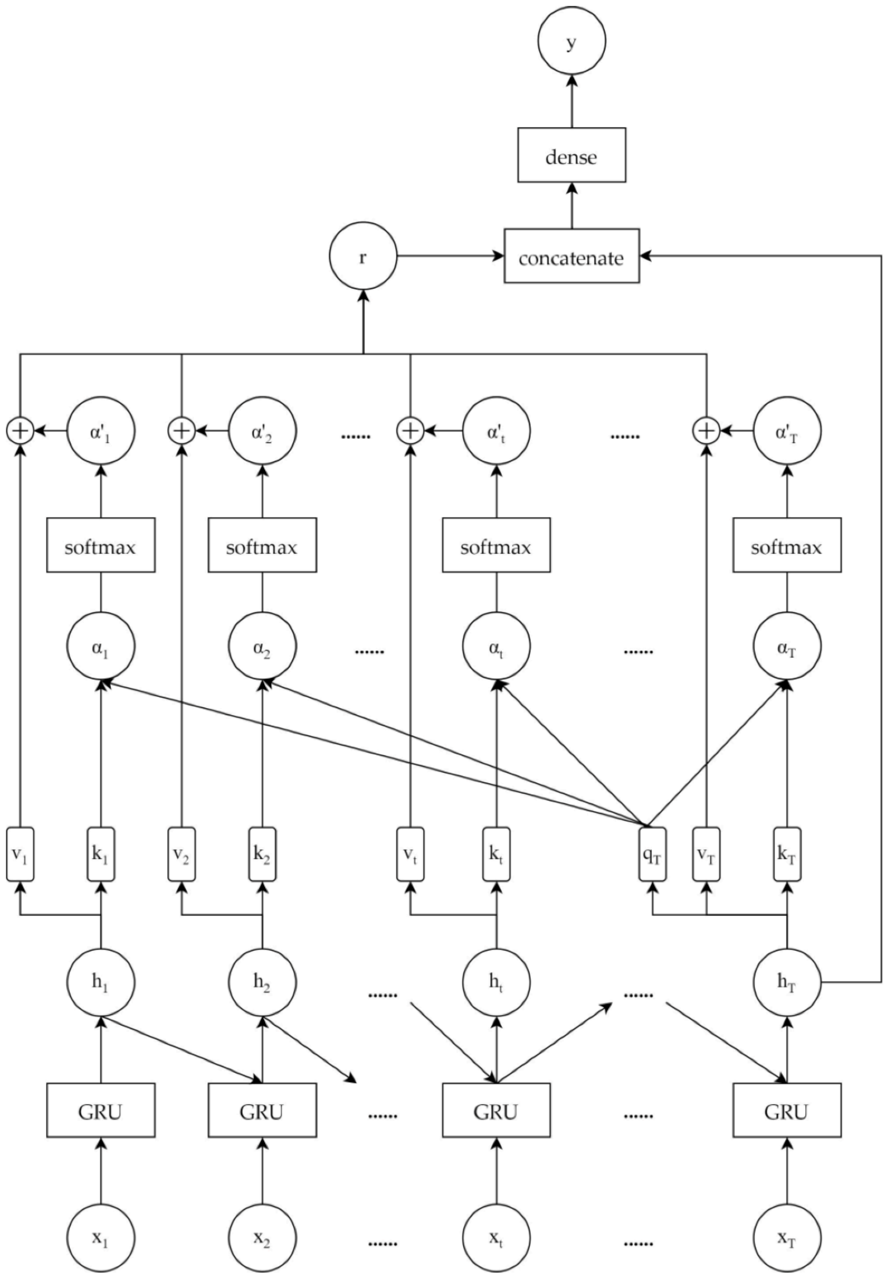


Fig 2.6 Attention mechanism In History Based Attention

To evaluate the accuracy of the model, the authors used data from US stock market and Chinese stock market. SP500 and CSI 300 from the span Jan2015 to December 2020 were taken as a input to the model. The results of the experiment showed that CEGH outperformed existing models in terms of accuracy. The authors also conducted a comparative analysis of CEGH with other models and found that CEGH had a higher accuracy than the other models.

Overall, the paper by Yijiao Liu, Xinghua Liu,Yuxin Zhang,Shuping Li provides a detailed analysis of the proposed hybrid model, CEGH, and its performance. The authors have shown that CEGH provides a more accurate forecast for intraday stock prices than existing models. The authors have also provided an in-depth analysis of the components of the model and their effect on the results. This paper provides a valuable insight into the use of hybrid models for intraday stock price forecasting and will be beneficial to investors in the stock market.

2.5.4. Frequency Decomposition Based GRU Transformer Neural Network:

In their paper, Chengyu Li and Guoqi Qian (2023) uses a frequency decomposition based GRU transformer neural network to attempt to predict stock prices. The methodology they use for this process involves decomposing the time series data into different frequency components, and then using a GRU transformer neural network to extract the features from the decomposed components. The dataset they use in their research includes stock market data from the S&P 500 index.

The main challenge they focused is dealing with high frequency data. For intra trading the data should be considered at minutes and second's level. When this data is fed to Lstm/Gru /Rnn networks they are unable to identify the fluctuations within the windows as the fluctuations are very less.
The input is decomposed into multiple Imf's and they are fed to the Lstm/Gru network. The output form the Gru networks is fed to the Transformer for the output. They used 100 stocks dataset from CSI-300.

The model is compared with baseline models DeepLOB, DeepAtt,MHF and it has achieved high R2 and less errors. For evaluation mean squared error, root mean squared error and Huber loss were used.

The authors highlight several limitations to their approach. For example, the accuracy of the model's predictions depends on the data it is given, and the input data can be noisy and incomplete. The authors also note that their model is limited in its ability to capture non-linear relationships between the inputs and the stocks' prices.

The authors suggest several areas of future work, including the development of more sophisticated models that can better capture non-linear relationships, and the use of a larger and more diverse dataset. They also suggest exploring the use of different architectures, such as long-short term memory networks, for the frequency decomposition task.

2.5.5. Stock Price Prediction Based on Temporal Fusion Transformer

This paper presents the temporal fusion transformer (TFT) model for stock price prediction. Stock price prediction is a challenging task due to the ever-changing economic market and the complexity of the stock market. The TFT model uses a combination of temporal features, such as historical stock prices and company financial data, to make predictions. In this paper, the authors evaluate the performance of the TFT model on a real-world dataset.

The authors use a dataset of daily stock prices and company financial data from the Shanghai Stock Exchange. The dataset includes more than 2000 stocks from over 10 years of trading history. The authors use a 70/30 split for training and testing. For the training set, the authors use five-fold cross-validation.

The authors use the TFT model for stock price prediction. The TFT model is a multi-layer temporal neural network that uses both temporal features and a learning-to-rank approach to make predictions. The authors compare the TFT model to several other models, including random forest, support vector machines, and recurrent neural networks.

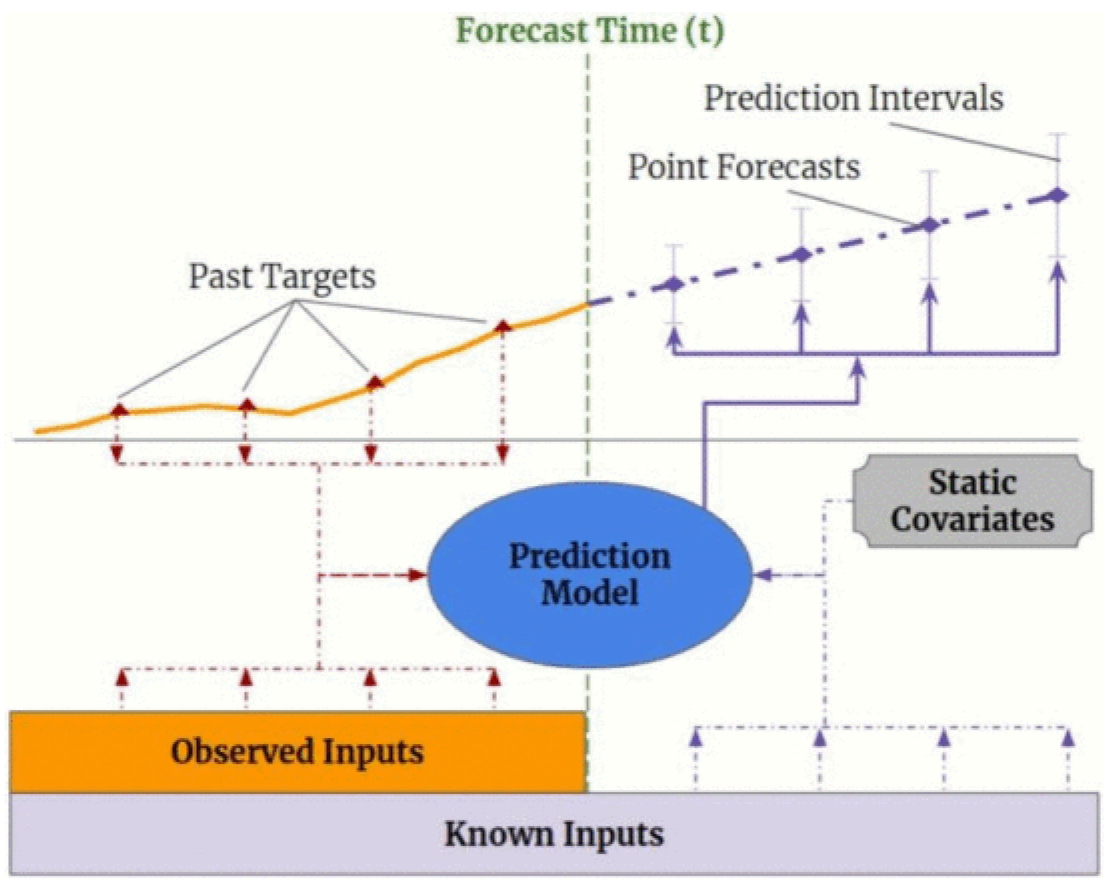


Fig 2.7 TFT Model Inputs

The authors find that the TFT model outperforms the other models in terms of accuracy and speed. The TFT model is able to make accurate predictions in a fraction of the time that it takes other models. However, there are some limitations to the TFT model. For instance, the TFT model does not take into account the economic or political context of the stock market. Additionally, the authors do not discuss the impact of data preprocessing on the performance of the TFT model.

The authors also identify some research gaps in the field of stock price prediction. For instance, they suggest that further research should focus on the analysis of different temporal features and their impact on the performance of the model. Additionally, they suggest that future research should focus on the development of models that can incorporate economic and political contexts into their predictions.

2.5.6. Transformer-based attention network for stock movement prediction

A literature review of Transformer-based Attention Network for Stock Movement Prediction by Qiuyue Zhang, Chao Qin, Yunfeng Zhang, Fangxun Bao, Caiming Zhang, and Peide Liu (2022) is presented. The paper introduces a new method for stock movement prediction based on a transformer-based attention network. The goal of the paper was to develop a method that could accurately predict the movement of stocks and to reduce the amount of manual input required by traditional methods.

The research gap that the paper highlights is the lack of methods for accurately predicting stock movement. The paper states that current methods are limited by the amount of manual input they require, and that more accurate methods are needed.

The methodology as shown in the below fig used by the authors to achieve this goal was to develop a transformer-based attention network for stock movement prediction. The text input is given to the multiheaded attention and daily trading prices are normalized for feature vector. Embedded vector generated from the text input is processed by the multiheaded attention and each head output is concatenated and given to the feed forward neural network. The fnn role is to capture the nonlinear relationship from the data. The fnn output is concatenated with the feature and given Lstm-attention based network to predict the price. In the experiment the window is taken 5 days i.e to predict the price it must look back past five days prices

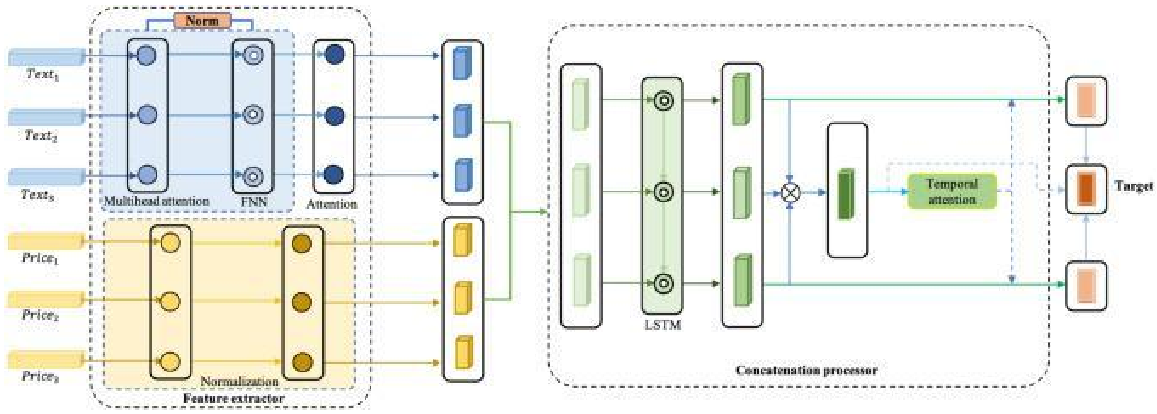


Fig 2.8. Transformer Based Attention Network

The dataset used to train the network consisted of stock data from the S&P500, NASDAQ, and NYSE markets. The data included information about stock prices, volume, and other market indicators. The dataset was split into training and testing sets, with the training set consisting of 80% of the data and the testing set consisting of the remaining 20%.

The paper does not discuss any limitations of the method presented. However, it is likely that the method is limited by the accuracy of the stock data used to train the network. Additionally, the method may be limited by the size of the dataset used.

The future scope for the paper is to further develop the method presented and to apply it to other datasets. Additionally, the authors suggest that the method could be used to develop algorithms for stock trading. They also recommend that the method could be used to monitor stock movements in real-time.

In conclusion, the paper introduces a new method for stock movement prediction based on a transformer-based attention network. The network was trained on a dataset of historical stock data and was able to accurately predict stock movements.

2.6. Discussion

Table2. Summarizes the literature review.

Author	Year	Title	Methodology	Datasets used	Evaluation Metrics and accuracy	Results and analysis
M. Reza Pahlaw	2021	Stock Price Foreca	RNN is used to predict the stock prices			Could be able to capture the

an, Edwin Riksakomara, and Raras Tyasnurita		st of Macro-Economic Factor Using Recurrent Neural Network				short-term relationships. The model Could not be able to identify the complex patterns
Y.Touzani, & Douzi	2021	LSTM and GRU based trading strategy adapted to the Moroccan market	Lstm and Gru is used to predict the stock prices	Two sources of data Dataset1: SP500 stocks extracted form yahoo finance. Dataset2: All stocks in CAC40 index collected form investing.com. The data covers from January 2010 to January 2019	Window period of five days was taken.MAPE, MSE,RMSE metrics was used for evaluation. For short-term prediction, the LSTM model has MAPE values for the COL and MDP stocks of 1.9% and 3.1%, respectively. Like this, the GRU model's MAPE values for medium-term prediction for the COL and MDP stocks are 2.1% and 4.3%, respectively.	The GRU model is utilised for medium-term price prediction, whilst the LSTM model is primarily designed for short-term price prediction. These models are trained using past price data, and they are then used to provide predictions that are considered when deciding how to proceed with the suggested trading plan.
Mondal Bhaska	2022	Machine Learni	Used different methodologies Decision	Dataset included macroeconomic conditions and	Gru performs better than other models.	The study says that by including

r, and Asimkiran Dandapat		ng and Deep Learning based Stock Market Prediction considering Covid-19 as a Feature	T,Radom Forest,svm,Lasso,Ridge,Bayesian Ridge,Gradient Boost,Lstm,Gru	financial market indices. The nifty 50return, Macroeconomic Values, Foreign Exchanges, Major indices	RMSE, R2, MAE values are used for evaluation	economic, financial and global data the machine and deep learning models can be able to predict the nifty 500 stock prices
Sabu, KS Rekha. MK	. 2022	A cooperative deep learning model for stock market prediction using deep autoencoder and sentiment analysis	Deep Autoencoder:Input layer,encoding deep belief network,compressed feature network,decoding deep belief network and output layer Sentiment analysis:VADER is used for sentiment analysis Time series analysis:LSTM/GRU was used for prediction	The top 25 news from 01/30/2000 to 01/07/2016 from Kaggle is used for sentimental analysis. Six vaules open,close,volume, high,low,adjusted close are collected from yahoo finance for three stocks HMC,ORCL,INTU	RMSE, MAE,MAPE, R2	For all the three stocks the DAE-LSTM and DAE-GRU achieved above 95% accuracy. This models accuracy better than grm and lstm ,also it includes sentimental analysis feature which is also a contribution factor for increasing accuracy
by Matin N. Ashtiani and Bijan Raahemi	2023	News-based Intelligent Prediction of Financial	The paper talks about various model used for technical and fundamental analysis.It talks about different nlp model used	Different data sources were used to collect news data of stocks such as Financial Times, The wall street,Journal,Bloomberg,Reuters,Forb	Common evaluation metrics used by the most papers referred in the article are precision,reca	The papers identifies few of the reasearch gaps like use finaniclor quarterly

		Markets Using Text Mining and Machine Learning: A Systematic Literature Review	mainly learning based and lexicon based	es and yahoo finance .Mainly webscarping and api are two ways to collect the data,but mostly api's are used to collect the data	ll,Mean squared Error,MAPE, RMSE	reports as they don't have noise when comapred to the social media or news articles .It also states the need of benchmark dataset to test state of art algorithhms ,size of the dataset ad data gathering limitations
Sridhar, Shashank, and Sowmya Sanagavarapu	2021	multi-head self-attention transformer (MAAT) for predicting the price of Dogecoin	Time vector and input sequence is concatenated and given to the multi headed attention transformer. The encoder part of the transformer captures the temporal dependencies and decoder part predict the future prices	The dataset contains the minute-by-minute trading data of dogecoin from 05 July 2019 to 28 April 2021	MSE, MAPE,R2,RMSE,Accuracy	The multiheaded self attention used ere is able to capture the hour-by-hour price variation. It can be able to capture the short term and long-term temporal dependencies
Ming-Che Lee, Jia-Wei Chang, Sheng-Cheng	2022	Applying attention-based BiLSTM and	Bilstm with attention is fed with technical indicators to predict the future price.The bilstm outputs	The dataset was collected from Taiwanese stock exchange from 2014/01/04 to 2021/05/31. The following technical	Precision, recall, accuracy, F1 measure. For STFT accuracy ranges from	For the STFT (Short-term Future Trend) example and the

Yeh, Tsorng-Lin Chia, Jie-Shan Liao, and Xu-Ming Chen		technical indicators in the design and performance analysis of stock trading strategies	the hidden states on which attention is applied and it is given to the dense layer to get the probability	indicators are used KD, RSI,BIAS,W%R,MACD	52.73% to 63.80% and for NDP accuracy ranges from 60.07% to 64.76%	NDP (Next Day Price) scenario, respectively, the average accuracy improvement in the experimental findings of the AttBiLSTM model is 3.35 percentage points and 4.52 percentage points, respectively. This indicates that the AttBiLSTM model performs noticeably better in terms of accuracy when compared to the basic LSTM model.
Yijiao Liu, Xinghua Liu,Yuxin Zhang, Shuping Li(2022	2022	A Hybrid Model Using CEEMD, Entropy, GRU, and History	The input sequences are given to the CEEMD where it decomposes it into different frequencies and is given to encoder to denoise the data. Then are	CSI300 from Chinese's stock market and sp500 from US market were taken. The sample period covers from Jan2015 to December 2020.20% of the	MAE,RMSE are used to evaluate the models prediction	The models are compared with the base line models. The MAE value of GRU-HA was 80% lower than GRU for

		Attention for Intraday Stock Market Forecasting	grouped into four groups and Gru- history attention is used to predict the prices	samples were taken for testing purpose		high frequency US Stock and for medium frequency it 35 % lower than GRU.These results indicate s that adding attention to the GRU increasing the performance of the model.The decomposition model mae is much 20% lower than Gru-HA model
Chengyu Li, Guoqi Qian.	2023	Frequency Decomposition Based GRU Transformer Neural Network	The input is decomposed into 4 series using Ceemd.Then it is fed to the LSTM layer followed by GRU layer and Transformer layer. The output of Transformer layer is concatenated with static time invariant features and fed to the linear	The limit order book of csi-300 dataset over 21 days collected for every 3s	Mean squared error,RMSE, Huber Loss	The model performance is compared with baseline models DeepLob,DeepAtt,MHF and it has lower r2,MAE,MSE

			layer to predict the output			
Hu, Xiaokang.	2021	Stock Price Prediction Based on Temporal Fusion Transformer	Temporal future transformer is a recent state of art transformer. It uses three inputs static metadata which is type of stock, time series data which stock prices and volumes, and known future time series which is exact forecasting date	The dataset contains the S&P500 companies from Jan4th 2010 to May31st 2021.It includes open, close,high,low,volume and adjusted close price as features	Mean absolute error and symmetric mean absolute percentage error	The model performance is compared to svr and lstm.when trained on small dataset is performance is not good but when trained big dataset the model performance is better than svr and lstm.Also it has features to include to time invariant features and it is a recent state of art algorithm
Qiuyue Zhang, Chao Qin, Yunfeng Zhang, Fangxun Bao, Caiming Zhang, Peide Liu.	2022	Transformer-based attention network for stock movement prediction	The text input is passed to transformer encoder and attention is applied. Then this output is concatenated with feature vector which contains daily trade prices. This passed to the lstm and attention mechanism to predict the prices	The dataset contains 88 stocks with the period covered from Jan 2014 to Jan 2016.It also covers the data from different datasets which covers the period from 2008 to 2019 from github	Accuracy and Mathews correlation coefficient	The performance of the model is compared with the base line models, TSDLA, HAN,CH-RNN,STOCKNET,ADV-LSTM,CapTE.CapTE scores highest among the models. The model exists

						more than 20% accuracy when compared to the four baseline models

The results of these studies show that hybrid models, such as the CEGH model by Liu et al. (2022) and the hybrid information mixing module by Choi et al. (2023), improved performance compared to traditional models. Transformer models, such as the temporal fusion transformer by Hu (2021) and the transformer-based attention network by Zhang et al. (2022), have also shown to be effective in predicting stock prices. These models are effective at capturing temporal dependencies between input data and stock prices. The studies also showed that including multiple data sources, such as macroeconomic factors (Pahlawan et al., 2021) and sentiment analysis (Parvatha et al., 2023), can improve accuracy. Additionally, including time-invariant data, such as technical indicators (Lee et al., 2022), can also improve accuracy. The temporal relationships between long and short terms can be captured using transformer models. These models are able to learn complex patterns from large datasets, and therefore can be used to identify seasonal patterns and weekly patterns. For example, the Block-Recurrent Transformer (BRT) model proposed by DeLesley Hutchins et al. (2022) uses recurrent layers to capture temporal relationships and transformer layers to capture long-term dependencies. Moreover, the Hybrid Information Mixing Module (HIMM) proposed by Choi et al. (2023) uses attention weights to identify the most important time series features and uses recurrent layers to capture temporal dynamics. Other transformer-based models such as Transformer-based Attention Network (TAN) (Qiuyue et al., 2022) and Temporal Fusion Transformer (TFT) (Hu, 2021) can also be used for stock market prediction. Additionally, Li et al. (2023) proposed a frequency decomposition-based model which combines GRU and transformer layers to capture both short-term and long-term temporal relationships. Finally, Touzani and Douzi (2021) proposed a hybrid model combining LSTM and GRU layers to adapt to the Moroccan stock market. All of these models demonstrate how temporal relationships can be captured using transformer models for stock market prediction.

2.7. Following research gaps are identified in the literature review.

Inclusion of Quarterly Reports: The research mentioned above uses news articles, social sentiment analysis, and daily trade prices as their primary data sources. However, there is still a research gap because quarterly reports were not used as an input. Quarterly reports offer useful financial data that have a big impact on stock prices. Such reports can improve the accuracy of predictive models and give a more thorough grasp of market dynamics.

Incorporating Static Time-Invariant Features: The studies above showed a strong emphasis on the use of time-variant information, such as sentiment analysis and previous price data. However, there is a significant study gap when it comes to accounting for static time-invariant variables like industry sector, market capitalization, or financial ratios. These attributes offer essential details about businesses and can be used to identify long-term trends and consumer behaviour. The ability of the models to predict outcomes can be increased by combining static and time-variant information.

Identifying Long-Term and Short-Term Temporal Dependencies: The research above highlighted d how well transformer models capture the temporal relationships between input data and stock prices. Both long-term and short-term temporal dependencies can be effectively handled by transformer models. By choosing models that can recognise both long-term and short-term temporal connections, this fills a research need. The results show that transformer-based models may capture these relationships and boost prediction accuracy, as can hybrid models that include recurrent and transformer layers.

2.8. Summary

This literature review explores the use of Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) models to forecast stock prices. It then delves into hybrid models that use attention mechanisms to capture long-term dependencies and outperform recurrent models. Attention mechanisms have been shown to be effective for capturing long-term dependencies, while recurrent networks are better suited for short-term dependencies. Many hybrid models first use LSTM, then add an attention mechanism to the hidden states for a more effective long-term forecast. For intraday stock trading or high-frequency data, frequency decomposition and entropy have been added to remove noise and improve accuracy. Additionally, decoding and encoding architectures have been used to remove noise from the data before feeding the data into recurrent networks, resulting in improved accuracy.

CHAPTER3: RESEARCH METHODOLOGY

3.1. Introduction:

The proposed methodology blends sentiment analysis with the Temporal Fusion Transformer (TFT) model using Finance BERT, a unique language model for financial sentiment analysis. The TFT model was chosen because it can effectively capture temporal dependencies, which are essential for stock market forecasting, in time series data. Contrarily, Finance BERT is designed specifically for sentiment analysis in the financial industry and is capable of successfully extracting sentiment from news articles.

The research uses comprehensive dataset of news articles, making sure it includes a variety of pertinent financial news sources. We will use methods like text cleaning, tokenization, and feature extraction to preprocess the news articles textual content. The sentiment indicated in the news articles will subsequently be examined using sentiment analysis techniques, such as Finance BERT or other appropriate models. We will receive sentiment scores from this step of the sentiment analysis process that will indicate whether a news report has a favourable, negative, or neutral sentiment.

The research also focuses on including time-invariant factors that are taken from a variety of sources, such as industry statistics, macroeconomic indicators, and company-specific financial measurements. The feature space will be enhanced by these time-invariant characteristics, which will also add more context that is constant across the dataset.

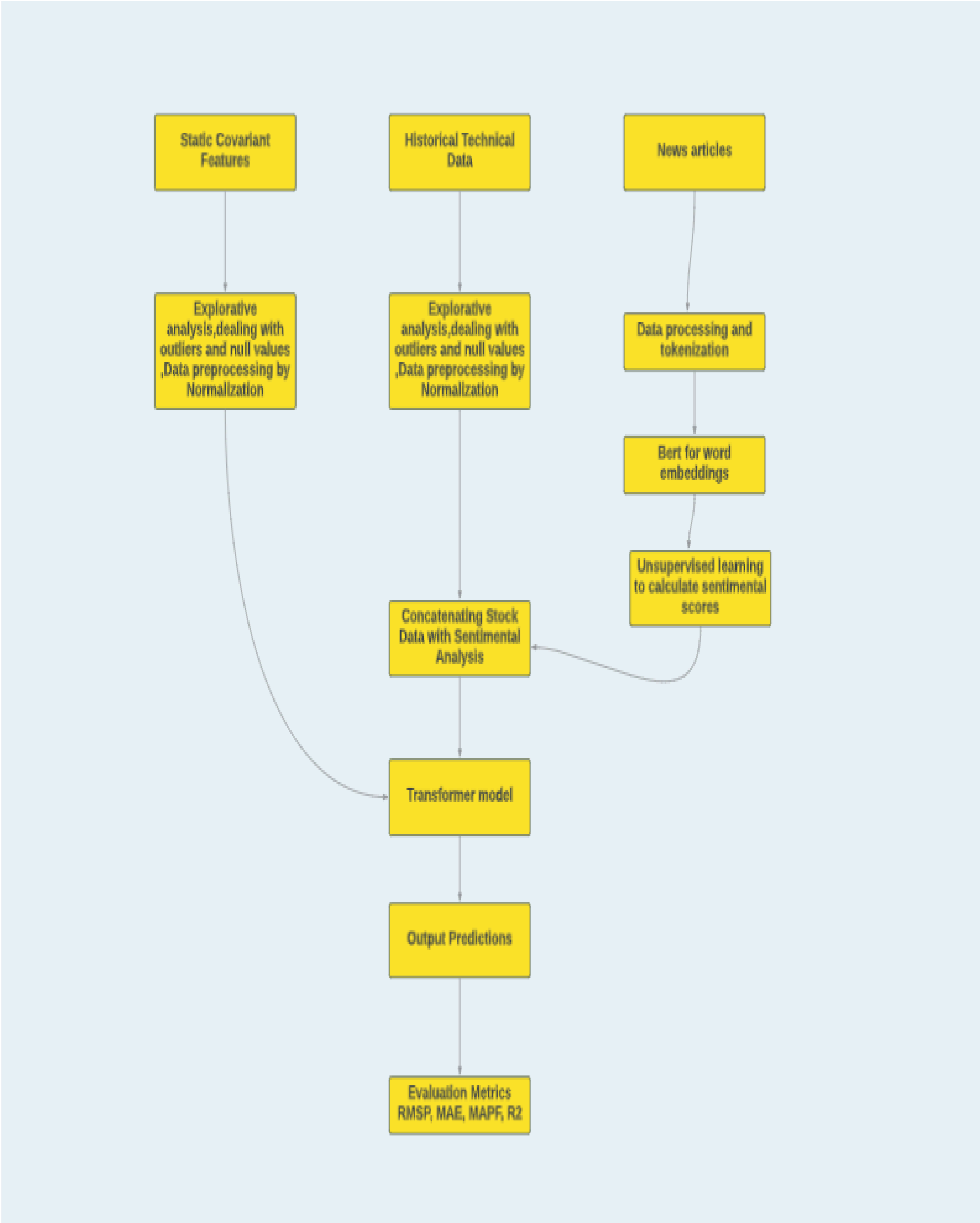
The time-variant elements, such as daily trade prices, volume, and technical indicators, will be combined with the sentiment ratings score from news items and the time-invariant parameters. The TFT model will get this combined feature vector as input, allowing it to account for both temporal dependencies and sentiment-driven market dynamics.

The performance of the model is compared with the baseline models like Lstm,Gru and attention based Lstm by including sentimental score and time invariant parameters and also by excluding them. Evaluation metrics such as precision, accuracy, f1 score, RMAE, recall scores will used access the model predicting capability comparing to the baseline models.

The findings of this study will help to understand the function of sentiment analysis on news articles and the addition of time-invariant characteristics in stock market forecasting more thoroughly. Insights from the findings will help investors, financial analysts, and academics make wise investment decisions by illuminating how these aspects can improve the performance of the TFT model.

In the sections that follow, we'll go over how the data was gathered, how sentiment analysis of news reports was done, how time-invariant parameters were incorporated, how the TFT model was built, how the experiment was set up, how the results were analysed, how the study was limited, and how future research might be pursued.

3.2. Methodology



Static Covariant Features
Historical Technical Data
News articles
Explorative analysis,dealing with outliers and null values ,Data preprocessing by Normalization
Explorative analysis,dealing with outliers and null values ,Data preprocessing by Normalization
Data processing and tokenization
Bert for word embeddings
Unsupervised learning to calculate sentimental scores
Concatenating Stock Data with Sentimental Analysis
Transformer model
Output Predictions
Evaluation Metrics RMSP, MAE, MAPF, R2

Fig 3.1. Project Methodology Flow chart

The below steps elaborate the methodology.

1)Historical Data Price:

Collect historical stock price data from yahoo finance Api. The features include stock closing price, opening price, high price, low price and volume.

2)static covariant features:

Time invariant features for a stock include industry of the stock, financial indicators or other external factors that does not change over time but influence the stock price.

3)News Article:

Collect News articles for the related stocks and process the data The tokens are passed into the Bert for word embeddings and then passed through the unsupervised learning to get the sentimental scores for the word embeddings.

4)Data Normalization:

This step involves normalizing the data by using min max scaling. This helps the model to converge faster.

5)Concatenation:

The obtained feature vector is concatenated with sentimental score and time invariant features. Then that input is fed into the TFT model

3.3. Dataset:

The Finhub API provides news articles, historical prices and company related data. It has premium and free services; under premium it provides access to various API endpoints. First step is to need to signup for the api key and then by using finhubclient we can be able to call the respective endpoints.

To collect the news articles and static invariant features Finhub API is used.

The following python code is used to collect the News articles from the finhub api

```
import finnhub
import pandas as pd
import requests
import time
# Set up API parameters
```

```
stock = 'AAPL'
from_date = '2022-01-01'
to_date = '2022-12-31'
api_key = "choo12hr01qjvijln0h0choo12hr01qjvijln0hg" # Replace with your Finnhub API key
finnhub_client = finnhub.Client(api_key="choo12hr01qjvijln0h0choo12hr01qjvijln0hg")
schema = {
    "category": str,
    "datetime": pd.Int64Dtype(),
    "headline": str,
    "id": pd.Int64Dtype(),
    "image": str,
    "related": str,
    "source": str,
    "summary": str,
    "url": str
}
i=14
least_date=pd.to_datetime('2022-12-12')
# Iterate over stocks and retrieve data
while least_date >= pd.to_datetime('2022-12-01'):
    # Make API request
    #url =
f'https://finnhub.io/api/v1/company-news?symbol={stock}&from={from_date}&to={to_date}&t
oken={api_key}'
    response = finnhub_client.company_news(stock, _from="2022-12-01",
to=least_date.strftime('%Y-%m-%d'))
    #data = response.json()

    # Convert data to DataFrame
    df = pd.DataFrame(response,columns=schema.keys()).astype(schema)
    df['datetime'] = pd.to_datetime(df['datetime'], unit='s')
    # Save DataFrame to CSV
    filename = f'{stock}_news_8_{i}.csv'
    df.to_csv(filename, index=False)
    print(f'Saved data for {stock} to {filename}')
    # Sort DataFrame by date
    i=i+1
    df_sorted = df.sort_values('datetime')
    # Find the least date
    least_date = df_sorted['datetime'].iloc[0]
    new_date = least_date - pd.DateOffset(days=1)
    least_date=pd.to_datetime(new_date)
    print(f"The least date minus one day is: {new_date}")
```

[redacted]

To collect the historical trading prices yahoo finance API is used. The following is the code in the python to collect the data from yahoo finance.

```
import yfinance as yf


# Define the ticker symbol for the stock

ticker = "AAPL"


# Set the start and end dates for the historical data

start_date = "2022-01-01"

end_date = "2022-12-31"


# Define the desired columns

columns = ["Open", "High", "Low", "Close", "Volume"]


# Retrieve the historical data from Yahoo Finance

data = yf.download(ticker, start=start_date, end=end_date, columns=columns)


# Print the retrieved data

print(data)
```

3.4. Temporal Fusion Transformer Architecture

The difficulty of handling a complicated mix of inputs, such as static covariates, knowing future inputs, and variant time series observed only in the past, is addressed. Traditional deep learning models for multi-horizon forecasting are frequently referred to as "black box" models because of their lack of transparency into the way in which they make use of the many inputs found in real-world scenarios. The goal of TFT, in contrast, is to offer understandable insights into the temporal dynamics of the data.

Recurrent Layers: Recurrent layers for local processing are incorporated into TFT, enabling the model to capture temporal interactions at various scales. Recurrent layers are good at identifying patterns and short-term dependencies in the data.

Interpretable Self-Attention Layers: TFT also makes use of self-attention layers, which provide the model the ability to recognise long-term dependencies and comprehend the connections between various time steps. The model can concentrate on pertinent information because of self-attention mechanisms that give relevance weights to various items in the input sequence.

The performance and interpretability of TFT are further improved by specialised components for feature selection and gating layers for suppressing extraneous components. These processes assist the model in choosing pertinent features and removing noise or unnecessary data, hence increasing the accuracy of its forecasts across a wide range of scenarios.

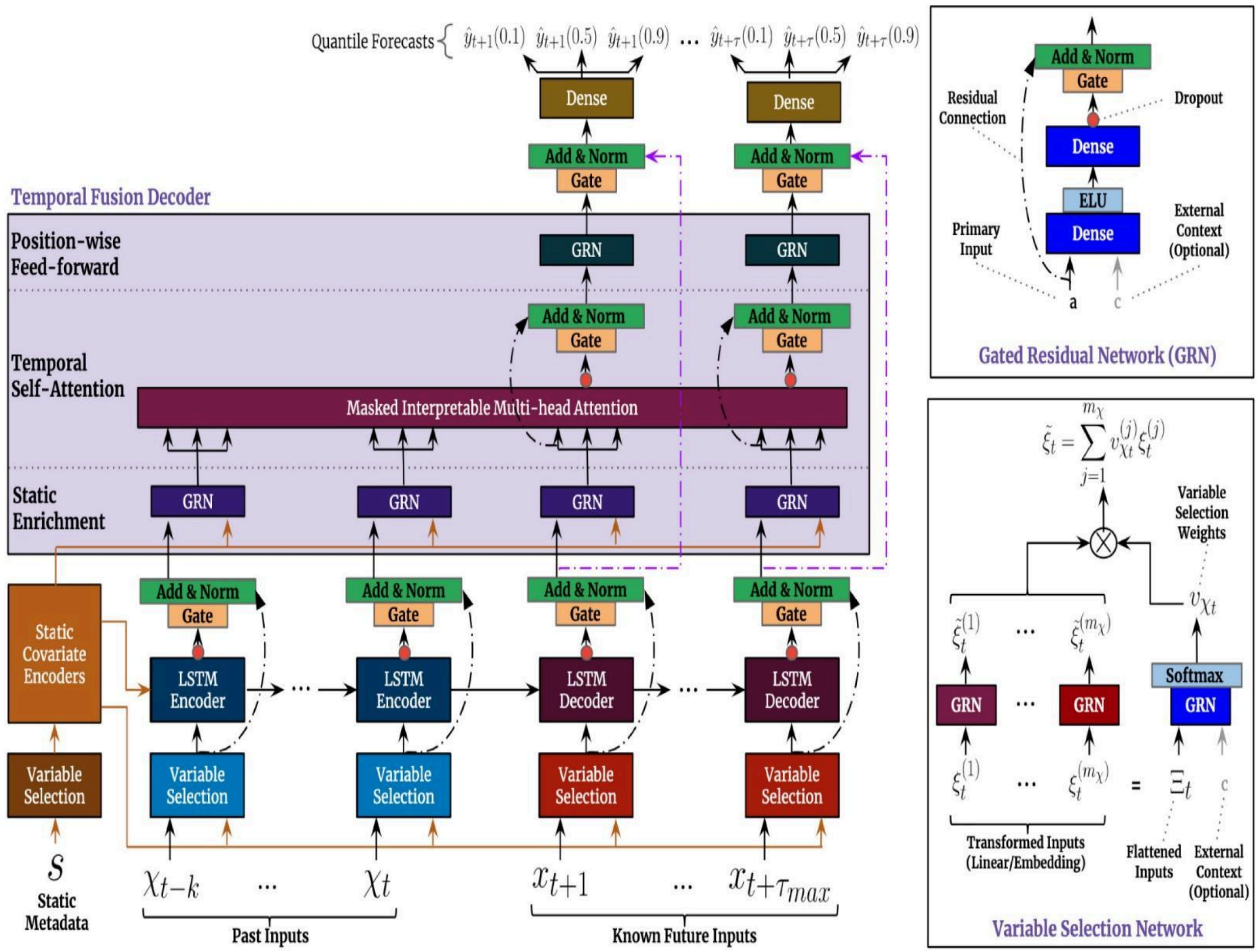


Fig: 3.2 Temporal Fusion Transformers Architecture

Multi-horizon forecasting: It is used to describe the forecast of important variables at several future time steps. Given that it gives users estimations of the target variable across the full forecast horizon, it is a significant problem in time series machine learning. This gives users the ability to choose wisely at various subsequent steps and optimise their activities, which can be useful in a variety of real-world applications.

Multiple data sources that can offer helpful information for developing precise predictions are frequently accessible in realistic multi-horizon forecasting scenarios.

Known Future Information: When making forecasts, this relates to knowledge about upcoming occurrences or influencing factors that is already available Earnings Release Dates, Dividend Payment Dates, Stock Splits or Reverse Splits, for instance, might be regarded as known future information in the context of retail forecasting.

Exogenous Time Series: These additional historical time series data have been seen before and are thought to influence the target variable. For instance, stock price data can be viewed as an exogenous time series that may have an impact on sales in the retail industry.

Static Metadata: Fixed or time-invariant properties or characteristics connected to the data are referred to as static metadata. This can include details like a stock market capitalization or stock domain characteristics. Forecasting can benefit from the contextual knowledge that static metadata can offer.

Utilising these diverse data sources efficiently and comprehending their interactions with one another is difficult when forecasting multi-horizon time series. The interconnections between many data sources could not be recognised beforehand, which makes the issue more challenging.

The Temporal Fusion Transformer (TFT) is introduced in the study as an alternative to current techniques. TFT blends interpretable self-attention layers for capturing long-term dependencies with recurrent layers for local processing. It can handle a variety of inputs, including time-varying inputs with known past and future values as well as static variables. To increase performance, TFT uses specialised components for feature selection and gating layers. Furthermore, TFT analyses attention patterns to offer perceptions of temporal dynamics while preserving cutting-edge performance on varied datasets.

3.4.1. Model architecture

It has a Gating mechanism, Variable selection networks, Static covariate encoders, Temporal processing, Prediction intervals.

3.4.1.1. Gating mechanisms:

The gated mechanism is a method used in neural networks to selectively emphasise or suppress components or layers in order to manage the flow of information. It enables the model to select the data that is pertinent to the task at hand in an adaptable manner. The Gated Linear Unit (GLU), a typical kind of gating mechanism, is commonly used. GLU determines the gating coefficients using a sigmoid activation function.

$$GLU\omega(\gamma) = \sigma(W4,\omega\ \gamma + b4,\omega) \odot (W5,\omega\ \gamma + b5,\omega)$$

The output of a preceding layer or component may be the input to the GLU in this equation, which is denoted by the symbol. W4, W5, b4, and b5 are bias vectors, stands for element-wise multiplication, and W4 and W5 are weight matrices.The first set $W4,\omega\ \gamma + b4,\omega$ decides how much input to supress or attenuate and second set $W5,\omega\ \gamma + b5,\omega$ determines how much gating should be allowed

$$GLU\omega(\eta 1) = LayerNorm(a + GLU\omega(\eta 1))$$

$\eta 1$ represents the intermediate output of the first equation.

a represents the primary input to the GRN.

Layer Norm denotes the layer normalization operation. This equation normalizes the output across the hidden dimensions.

The above two equations represent the gating mechanism which suppress the input and gives the model the ability to adaptively modify the flow of information throughout the network.

3.4.1.2. Variable selection networks:

The processed feature vectors (t) and an outside context vector (cs) are both inputs to the variable selection network. The context vector, which is derived from a static covariate encoder, contains details regarding the static variables.The Gated Residual Network is applied to the concatenated input . For each input variable to be considered relevant, GRNvχ employs a combination of suppression and gating procedures. The softmax layer, which normalises the outputs and creates a vector of variable selection weights known as Vt, is used to process the, GRNvχ output after that. The importance of each input variable at the specified time step is indicated by these weights.

$$v\chi t = \text{Softmax}(GRNv\chi(\Xi t, cs))$$

$$\tilde{\xi} t = \sum j=1m\chi\ v(j)\chi t\ \tilde{\xi}(j)t$$

The resulting vector, represented as t, represents the features that have been chosen and are thought to be most important for the prediction at time step t. The performance and interpretability of the model can be enhanced by employing the variable selection process, which allows TFT to concentrate on the most relevant input variables and eliminate any unwanted or noisy inputs.

3.4.1.3. Static covariate encoders:

The static metadata is incorporated by using separated encoders which four context vectors cc,ch,ce and cs.

Context for Temporal Variable Selection (cs):

By using the GRNcs on the result of the static variable selection network, the context vector cs is produced.By using this context vector specifically for temporal variable selection, the model is able to assess the value and contribution of various variables to the task of prediction.

Contexts for Local Processing of Temporal Features (cc, ch):

In the local processing of temporal data, the context vectors cc and ch play a role.

These context vectors are wired into precise places in the temporal fusion decoder to offer additional knowledge and direction for efficiently processing temporal characteristics.

Enrichment of Temporal Features with Static Information (ce):

The static information is added to the temporal features through the context vector ce.It is created by using the GRNce on the output ζ of the static variable selection network.The model can benefit from the added knowledge that the static metadata provides by adding static data into the temporal features.

The interpretable multi-head attention, calculates the relevance of various input components by taking into account how they relate to the keys and queries.It employs multiple attention heads to capture multiple aspects of the input. All of the attention weights from heads are merged to form a single set of attention weights, which is then applied to the values.This modified attention mechanism improves interpretability.

The scaled dot-product attention can be expressed as:

A(Q, K) = Softmax(QKT / √dattn)

The multi-head attention can be represented as:

MultiHead(Q, K, V) = [H1, ..., HmH]WH Hh = Attention(QW(h)Q, KW(h)K, VW(h)V)

InterpretableMultiHead(Q, K, V) = H̃W̃H H̃ = Ã(Q, K) V WV Ã(Q, K) = 1/mH ∑h=1 A(QW(h)Q, KW(h)K)

3.4.1.4 Temporal fusion decoder:

Locality enhancement with sequence-to-sequence layer

Time series data are targeted by the locality enhancement layer in TFT in order to identify regional patterns and context. A sequence-to-sequence layer with LSTM encoding and decoding is used to achieve this. Based on the values in their immediate vicinity, this layer aids the model in identifying noteworthy sites. In order to maintain important features and enhance information flow, it also uses static metadata and a controlled skip connection. Overall, the locality enhancement layer improves the model's capacity to recognise significant temporal patterns in the data and comprehend local interactions.

References

Zheng, Y., Li, Y., Zhang, L., & Huang, H. (2020). Stock Market Prediction Using Optimized LSTM and GRU Models. arXiv preprint arXiv:2005.04707.

Touzani, Y. & Douzi, K. An LSTM and GRU based trading strategy adapted to the Moroccan market. J. Big Data 8, 126 (2021). https://doi.org/10.1186/s40537-021-00289-9.

Li, X., Li, X., Zhang, Y., & Wang, H. (2019). Stock Market Prediction Based on Optimized LSTM and GRU Models. In 2019 15th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD) (pp. 749-753). IEEE.

Alejandro Lopez-Lira, Yuehua Tang. 2023. "Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models."

Burra, Lakshmi Ramani, Bhanu Prasanna Koppolu, Baduganti Deva Karthik, Bolem Lasya Naga Sai Priya, Allam Prasanthi, and Praveen Tumuluru. 2023. "Stock Price Prediction using Zero-Shot Sentiment Classification."

Chengyu Li, Guoqi Qian.2023. "Stock Price Prediction Using a Frequency Decomposition Based GRU Transformer Neural Network."

Choi, Jooweon, Shiyong Yoo, Xiao Zhou, and Youngbin Kim. 2023. "Hybrid Information Mixing Module for Stock Movement Prediction."

DeLesley Hutchins, Imanol Schlag, Yuhuai Wu, Ethan Dyer, Behnam Neyshabur. 2022. "Block-Recurrent Transformers."

Fernando Chague, Rodrigo De-Losso,Bruno Giovannetti. 2019. "Day Trading for a Living?"
Heliyon. 2021. "The performance of retail investors, trading intensity and time in the market: evidence from an emerging stock market."

Hu, Xiaokang. 2021. "Stock Price Prediction Based on Temporal Fusion Transformer."
Jitender Kumar, Neha Prince. 2022. "Overconfidence bias in the Indian stock market in diverse market situations: an empirical study."

K, Karthik, Ranjithkumar V, Sanjay Kiran K P, and Santhosh Kumar P S. 2023. "A Survey of Price Prediction using Deep Learning Classifier for Multiple Stock Datasets."

Khalid Alkhatib, Huthaifa Khazaleh,Hamzah Ali Alkhazaleh,Anas Ratib Alsoud,Laith Abualigah. 2022. “A New Stock Price Forecasting Method Using Active Deep Learning Approach.”

M. Reza Pahlawan, Edwin Riksakomara,Raras Tyasnurita. 2021. “Stock price forecast of macro-economic factor using recurrent neural network.”

Matin N. Ashtiani, Bijan Raahemi. 2023. “News-based intelligent prediction of financial markets using text mining and machine learning: A systematic literature review.”

Ming-Che Lee, Jia-Wei Chang, Sheng-Cheng Yeh, Tsorng-Lin Chia, Jie-Shan Liao,Xu-Ming Chen. 2022. “Applying attention-based BiLSTM and technical indicators in the design and performance analysis of stock trading strategies.”

Mondal, Bhaskar, and Asimkiran Dandapat. 2022. “Machine Learning and Deep Learning based Stock Market Prediction considering Covid-19 as a Feature.”

Parvatha, Leela Satish, Devu Naga Veera Tarun, Mentem Yeswanth, and Jonnalagadda. Surya Kiran. 2023. “Stock Market Prediction Using Sentiment Analysis and Incremental Clustering Approaches.”

Qiuyue Zhang, Chao Qin, Yunfeng Zhang, Fangxun Bao, Caiming Zhang, Peide Liu. 2022. “Transformer-based attention network for stock movement prediction.”

Reddy, K. Ritwik, B. Tarun Kumar, V. Rohit Ganesh, Polisetty Swetha, and Prakash Kumar Sarangi. 2022. “Stock Market Prediction Using Recurrent Neural Network.”
Sabu, KS Rekha.MK. 2022. “A cooperative deep learning model for stock market prediction using deep autoencoder and sentiment analysis.”

Sridhar, Sashank, and Sowmya Sanagavarapu. 2021. “Multi-Head Self-Attention Transformer for Dogecoin Price Prediction.”

Liu, Y., Liu, X., Zhang, Y., & Li, S. (2022). CEGH: A Hybrid Model Using CEEMD, Entropy, GRU, and History Attention for Intraday Stock Market Forecasting.