**APPROVAL PAGE**

**TITLE:** TITLE OF PROJECT PAPER

**AUTHOR:**

The undersigned certify that the above candidate has fulfilled the condition of the project paper in partial fulfilment for the degree of Master of Information Technology.

**SUPERVISOR:**

Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Name:

Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**ENDORSED BY:**

Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Name:

Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

STOCK INDEX’S TRADING BY MACHINE LEARNING ANALYSIS

XXXXXX

PROJECT REPORT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF INFORMATION TECHNOLOGY

UNITAR INTERNATIONAL UNIVERSITY

MONTH AND YEAR OF SUBMISSION

# DECLARATION

I hereby declare that the work in this project is my own except for quotations and citations which have been duly acknowledged.

DATE \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

XXXXXXXXXXXX

# ABSTRACT

TBU

The report should start with a short abstract stating clearly the objectives, scope and results of the work presented. It should be in only one paragraph and contain no more than 400 words.

P1 >> intro, what is the problem, purpose P2 >> why problem hard? What literature review mentioned? What are the obstacles? P3>> how you solve it, what you do? And results&finding P4 >> impact and effect of knowing this result (conclusion)

Aims: have direction of study >> collecting existing work, point out things haven’t study/ weakness/ improvement, decide your scope/ question to proceed >> collecting data & experiment method (save all results too) >> interpret result >> analyse & discuss

Section {intro, body, conclusion}, paragraph {[linked before] introduce one idea, describe, conclude idea [linked next] }

Use active phase for literature review, Past tense for results, no first-person word,

# ACKNOWLEDGEMENT

First of all, I would like to express my sincere gratitude to the many individuals who have supported and guided me throughout the journey of researching and writing this thesis on machine learning prediction of a stock index.

First and foremost, I would like to extend my heartfelt appreciation to my thesis supervisor, Dr. Farhad. His patient support and continuous encouragement have been instrumental in shaping this research. His guidance, constructive feedback, and insightful discussions have not only enriched the quality of this thesis but also significantly contributed to my personal and academic growth. I am truly fortunate to have had Dr Abdulaziz Yahya Al-Nahari as my mentor, and I am deeply thankful for their mentorship.

I would also like to thank my university’s department, Faculty of Business & Technology for providing me with the necessary resources, access to data, and a conducive research environment.

After that, I want to thank my friends and colleagues too. They supported me, talked to me about my work, and shared their experiences. It was great to have them with me during this challenging but exciting time.

Lastly, I want to thank my family. They were always there for me, supporting me and understanding why I was so busy. Their love and support meant everything to me.

# TABLE OF CONTENTS

[DECLARATION I](#_Toc180112283)

[ABSTRACT II](#_Toc180112284)

[ACKNOWLEDGEMENT III](#_Toc180112285)

[TABLE OF CONTENTS IV](#_Toc180112286)

[LIST OF FIGURES VIII](#_Toc180112287)

[LIST OF TABLES IX](#_Toc180112288)

[LIST OF APPENDICES X](#_Toc180112289)

[1. INTRODUCTION 1](#_Toc180112290)

[1.1 Background of the Study. 1](#_Toc180112291)

[1.2 Problem Statement 2](#_Toc180112292)

[1.3 Research Objectives 2](#_Toc180112293)

[1.4 Research Questions 3](#_Toc180112294)

[1.5 Significance of the Study 4](#_Toc180112295)

[1.6 Scope and Limitations 4](#_Toc180112296)

[1.7 Thesis Structure 5](#_Toc180112297)

[2. LITERATURE REVIEW 7](#_Toc180112298)

[2.1 Overview 7](#_Toc180112299)

[2.2 Stock 8](#_Toc180112303)

[2.2.1 Stock fundamental analysis 8](#_Toc180112304)

[2.2.2 Constraint of Individual stock price prediction 9](#_Toc180112305)

[2.3 Stock index 10](#_Toc180112306)

[2.4 Macroeconomics 11](#_Toc180112307)

[2.4.1 Macroeconomics & financial market 12](#_Toc180112308)

[2.4.2 Economics cycle 12](#_Toc180112309)

[2.5 Technical analysis of trend prediction 12](#_Toc180112310)

[2.6 Fundamental Analysis and Technical Analysis Review 13](#_Toc180112311)

[2.6.1 Candlesticks and OLHC 14](#_Toc180112312)

[2.6.2 Indicator 14](#_Toc180112313)

[2.6.3 Price Action Analysis and Cycle 15](#_Toc180112314)

[2.6.4 Impulsive and correction move 17](#_Toc180112315)

[2.7 Machine learning introduction 18](#_Toc180112316)

[2.8 Supervised learning 19](#_Toc180112317)

[2.9 Unsupervised learning 19](#_Toc180112318)

[2.10 Regression 20](#_Toc180112319)

[2.11 Semi supervised learning 21](#_Toc180112320)

[2.12 Reinforcement learning 22](#_Toc180112321)

[2.13 Convolutional Neural network (CNN) 23](#_Toc180112322)

[2.14 Deep learning 24](#_Toc180112323)

[2.15 Deep neural network (DNN) 24](#_Toc180112324)

[2.16 Recurrent neural network (RNN) 25](#_Toc180112325)

[2.17 Support Vector Machine 26](#_Toc180112326)

[2.18 Long short-term memory network (LSTM) 26](#_Toc180112327)

[2.19 Conclusion 27](#_Toc180112328)

[3. RESEARCH METHODOLOGY 30](#_Toc180112329)

[3.1 Overview 30](#_Toc180112330)

[3.2 Dataset Description 31](#_Toc180112331)

[3.3 Data Preprocessing 31](#_Toc180112332)

[3.4 Model Development 32](#_Toc180112333)

[3.5 Evaluation Metrics 33](#_Toc180112334)

[3.6 Implementation Tools 34](#_Toc180112335)

[3.7 Limitations 34](#_Toc180112336)

[4. SYSTEM DESIGN 34](#_Toc180112337)

[4.1 Overview 34](#_Toc180112338)

[4.2 System Architecture 35](#_Toc180112339)

[4.3 Data Collection and Preprocessing 36](#_Toc180112340)

[4.4 Feature Engineering 38](#_Toc180112341)

[4.5 Model Design and Training 38](#_Toc180112342)

[4.6 Prediction Workflow 39](#_Toc180112343)

[4.7 Technical Specifications 40](#_Toc180112344)

[5. RESULTS AND DISCUSSION 40](#_Toc180112345)

[5.1 Dataset Overview 40](#_Toc180112346)

[5.2 Model Performance Metrics 41](#_Toc180112347)

[5.3 Visual Analysis of Predictions 42](#_Toc180112348)

[5.4 Feature Importance 49](#_Toc180112349)

[5.5 Error Analysis 50](#_Toc180112350)

[5.6 Model Comparison 50](#_Toc180112351)

[6. Conclusion and Future Work 51](#_Toc180112352)

[6.1 Summary of Findings 51](#_Toc180112353)

[6.2 Implications for Practice 51](#_Toc180112354)

[6.3 Recommendations 52](#_Toc180112355)

[6.4 Future Research Directions 52](#_Toc180112356)

[6.5 Conclusion 53](#_Toc180112357)

[7. FORMATTING GUIDELINE 54](#_Toc180112358)

[7.1 Overview 54](#_Toc180112359)

[7.2 Document Layout 55](#_Toc180112360)

[7.3 Font size 55](#_Toc180112361)

[7.4 Spacing 55](#_Toc180112362)

[7.5 Page Numbering 55](#_Toc180112363)

[7.6 Heading 55](#_Toc180112364)

[7.6.1 Table 55](#_Toc180112365)

[7.6.2 Figure 56](#_Toc180112366)

[7.7 Printing 57](#_Toc180112367)

[7.8 Quoted Material 58](#_Toc180112368)

[REFERENCES 59](#_Toc180112369)

[APPENDIX 63](#_Toc180112370)

# LIST OF FIGURES

[Figure 7.1 Sample graph 9](#_Toc295896087)

[Figure 7.2 Sample graph 2 9](#_Toc295896088)

# LIST OF TABLES

[Table 7.1. Font sizes of headings 8](#_Toc295896098)

# LIST OF APPENDICES

# INTRODUCTION

## Background of the Study.

The stock market is one of the most dynamic and volatile financial markets, and predicting price movements has long been a critical area of interest for traders, investors, and financial analysts. The ability to accurately forecast stock prices can lead to significant financial gains, especially in short-term trading. Traditionally, traders have relied on technical analysis and fundamental analysis to make informed decisions about when to buy or sell stocks. However, with the rise of machine learning and artificial intelligence, new methods are being explored to enhance the accuracy and efficiency of stock price predictions.

One promising approach is the use of machine learning models, particularly deep learning architectures, to analyze historical price data and technical indicators. Deep learning models, such as Convolutional Neural Networks (CNNs), have shown remarkable success in fields like image recognition and natural language processing. Their ability to automatically extract patterns from complex data makes them well-suited for stock price prediction, where price movements are often influenced by intricate and interrelated factors.

In this study, we aim to develop a stock price prediction system using CNNs to forecast buy signals in the stock market. The focus is on short-term price prediction, specifically for intervals of 15 minutes, to assist traders in making quick and informed trading decisions. By leveraging technical indicators like moving averages, RSI (Relative Strength Index), and Bollinger Bands, combined with the power of deep learning, the system aims to provide actionable insights for traders.

## Problem Statement

Stock price prediction is a challenging task due to the inherent complexity and volatility of financial markets. Prices are influenced by a wide range of factors, including economic indicators, company performance, geopolitical events, and investor sentiment. Additionally, stock prices exhibit non-linear patterns and are subject to sudden, unpredictable changes, making accurate prediction difficult.

For short-term traders, the ability to identify buy signals—indications of when to purchase a stock before an upward price movement—can be particularly valuable. However, traditional technical analysis methods often fall short in capturing the complex relationships between price, volume, and other market indicators. Traders typically rely on heuristic rules or visual patterns, which can be subjective and prone to errors, especially in fast-moving markets.

This research seeks to address these challenges by developing a machine learning-based system that predicts buy signals using deep learning. Specifically, the study focuses on designing a CNN model that processes historical stock data and technical indicators to forecast buying opportunities in real-time.

## Research Objectives

The primary objective of this study is to design and implement a stock price prediction system capable of accurately forecasting buy signals for short-term trading. The system should be able to process historical stock data, compute relevant technical indicators, and predict potential upward price movements based on this data.

The specific objectives of the research are as follows:

1. **To collect and preprocess historical stock price data:** This includes gathering data from reliable financial sources and ensuring it is cleaned, normalized, and structured for model training.
2. **To design a Convolutional Neural Network (CNN) model for buy signal prediction:** The model will be trained to identify patterns in stock prices and technical indicators, with the goal of predicting when a stock is likely to experience an upward price movement.
3. **To evaluate the model's performance using real-world stock data:** The accuracy, precision, recall, and F1 score of the model will be measured to assess its effectiveness in predicting buy signals.
4. **To provide practical insights for traders:** The system should generate buy signal predictions that can be used by traders to make informed decisions. Visualizations of predicted buy signals alongside actual stock price movements will be included.
5. **To compare the performance of the CNN model with a traditional machine learning approach:** A baseline model using logistic regression will be implemented to compare the deep learning model's performance with a more conventional method.

## Research Questions

This study aims to answer the following key research questions:

1. Can a CNN-based model effectively predict buy signals in the stock market for short-term trading?
2. How do technical indicators, such as moving averages and RSI, contribute to the accuracy of buy signal predictions?
3. What is the comparative performance of a deep learning model versus a traditional machine learning model in predicting buy signals?
4. How well does the model generalize to different market conditions, including periods of high volatility?

## Significance of the Study

The significance of this research lies in its potential to provide traders with a reliable and automated tool for short-term stock price prediction. By integrating machine learning with traditional technical analysis, the system can offer more precise and data-driven insights, reducing the reliance on subjective judgment and improving trading outcomes.

For traders, particularly those engaged in high-frequency trading or short-term strategies, the ability to predict buy signals in real-time is crucial. An accurate and efficient prediction system can help traders capitalize on brief market movements, ultimately leading to higher returns. Additionally, by automating the prediction process, traders can reduce the time and effort required for manual analysis.

From an academic perspective, this study contributes to the growing body of research on the application of deep learning in financial markets. While much of the existing research focuses on long-term price prediction or market trends, this study explores the less-explored area of short-term buy signal prediction. The findings could inform future research on the use of machine learning in finance and lead to further advancements in stock market prediction systems.

## Scope and Limitations

The scope of this study is limited to the prediction of buy signals for short-term trading (15-minute intervals). The system focuses on technical analysis and does not incorporate fundamental analysis, news, or other external factors that may influence stock prices. The model is trained and tested using historical stock price data and technical indicators, and its performance is evaluated based on a limited set of stocks.

The main limitations of the study include:

* **Limited data sources:** The system relies on historical stock price data and technical indicators, which may not capture all relevant factors influencing stock prices.
* **Market volatility:** The model may struggle to perform well during periods of extreme volatility, where stock prices experience sudden and unpredictable changes.
* **Overfitting risk:** Despite efforts to prevent overfitting, the CNN model may still memorize specific patterns in the training data, leading to reduced generalization in unseen data.

## Thesis Structure

The structure of this thesis is as follows:

**Chapter 1:** Introduction (providing an overview of the study, problem statement, objectives, and research questions)

**Chapter 2:** Literature Review (examining previous research on stock price prediction, technical analysis, and machine learning methods)

**Chapter 3:** Methodology (describing the dataset, technical indicators, model architecture, and training process)

**Chapter 4:** System Design (detailing the design and architecture of the stock prediction system)

**Chapter 5:** Results and Discussion (presenting the evaluation of the model’s performance, including visual and quantitative analysis)

**Chapter 6:** Conclusion and Future Work (summarizing the findings, implications, and potential areas for further research)

# LITERATURE REVIEW

## Overview

This chapter surveys previous or current research relevant to the project topic. They are to be written to provide the research background and highlight the significant of the research. {{wide knowledge and results covered, idea supported by reference, relation/strength/constraint of these info., narrowed down scope and understand leftover things to study}}

* Discuss field being studied
* from old to new, from general to more specific
* Finding related information and study
* describe how its’ results help or related to your topic, with strength, weakness, constraint, support , fact}, popular opinion} “we think, we feel, we known …”, with reference, existing variable, possible variable, limitation, suggestion, idea
* relate these to your title and conclude at each session

This chapter will discuss about literature review made on

2.1 to 2.5 will discuss about stock market and factor affecting bull and bear of market, 2.6 to 2.10 will discuss about various machine learning model and algorithm. All these study and finding are concluded at 2.12



## Stock

Stocks represent units of ownership in a company that is traded publicly, making them accessible to anyone. When stocks are purchased, the buyer becomes a shareholder and gains the ability to vote on specific company decisions.

Stocks can be purchased in various ways, including through employee stock option plans, options trading, and traditional purchases facilitated by a broker. Options trading allows investors to buy a contract that grants them the right to purchase or sell a stock at a predetermined price in the future. Traditional stock purchases are conducted through a broker, which can be done either online or in-person.

If the company’s share price increase, the shareholder can make profit when selling their stocks, else, they will have financial loss.

### Stock fundamental analysis

Fundamental analysis is a method of evaluating a company's value and prospects. It considers qualitative factors, such as the company's business model, competitive advantage, management, corporate governance, and industry, as well as quantitative factors, such as financial statements like the balance sheet, income statement, and cash flow statement. Financial ratios drawn from data on corporate financial statements are used to make inferences about a company's value and prospects.

By evaluating both qualitative and quantitative factors, fundamental analysis provides a comprehensive understanding of a company's potential for growth and profitability. It is a useful tool for investors to determine whether a company's stock is undervalued or overvalued, and whether it is a good investment opportunity. (*Fundamental Analysis: Principles, Types, and How to Use It*, n.d.). Some of the effective mathematical indicator of company value that commonly refer by investor are Earnings per Share (EPS), Price-to-Earnings Ratio (P/E), Return on Equity (ROE), Price-to-Book Ratio (P/B), Price-to-Sales Ratio (P/S), Dividend Yield Ratio (DYR) and Projected Earnings Growth (PEG). These indicator do provide strong insight to diagnose company’s health and growth, hence suitable to be used to filter good company for investing in long term(Khairi et al., 2019), but it is less effective in predicting short term trading direction (Eiamkanitchat et al., 2017)

There is numerous factor other that company internal performance that can affect short term and long-term price, such as macroeconomics performance of counter, GDP, CPI, PPI, NFP, interest rate, industry and sector performance, market sentiment, predictable and non-predictable global events. (Matt Krantz, 2016) highlighted that fundamental analyst often express reservations about the approach of technical analysts. Fundamental analysts are primarily focused on factors like company performance and financial metrics, while technical analysts tend to place more emphasis on market momentum and dynamics. This divergence in perspective can sometimes lead to a phenomenon known as "groupthink," where market sentiment and trends influence prices to move further away from what fundamental analysts believe to be the true intrinsic value of an asset.

### Constraint of Individual stock price prediction

Efficient Market Hypothesis (EMH) is a theory in finance that suggests that financial markets are extremely efficient in processing and reflecting all available information about stocks, bonds, or other financial assets. It means that at any given time, the prices of these assets fully incorporate all known information, making it nearly impossible to consistently achieve above-average returns by trading based on publicly available information.

(Burton Gordon Malkiel, 2007) explains the EMH, which suggests that stock prices follow a random walk pattern, meaning they move in an unpredictable manner. He argues that the prices of individual stocks quickly reflect all available information, making it extremely difficult for investors to consistently predict their movements due to countless factors, including country and global macroeconomics, economic data, company news, even unpredictable events, can influence stock prices.

(Burton Gordon Malkiel, 2007) mentioned the challenges that investors face when trying to pick individual stocks, given the multitude of factors at play. He mentioned even professional money managers struggle to outperform the broader market consistently.

Due to constraint of individual stock, he suggests on Index investment, highlights how investing in broad market indices, like the S&P 500, can provide stability and a more predictable long-term return. These indices offer diversification, spreading risk across many companies, which can help mitigate the impact of poor-performing individual stocks.(Strader et al., 2020)

## Stock index

Stock index is a numerical representation of the performance of a group of selected stocks from a particular financial market or sector. It is a benchmark of how a specific segment of the stock market is performing. Stock indices are used to gauge the overall health and direction of the market, track changes in stock prices, and serve as a reference point for investors and financial professionals. Example of well-known stock indices include S&P 500 (United States), Dow Jones Industrial Average (DJIA or Dow 30, United States), FTSE 100 (United Kingdom), and Nikkei 225 (Japan), etc. Each of these indices represents a different aspect of its respective market and provides valuable information for investors and financial professionals.

(Burton Gordon Malkiel, 2007) describe the advantages of stock index as shown below:

1. Lower Costs of management fee and expenses for index investment than managed fund, with lower cost saving over time, it can increase net return of investment.
2. Good diversification in spreading risk across various sector. Hence poor performance of individual stock due to unpredictable event, poor performance of particular sector within a short period would not aggressively pull down your investment if you invest on index. Furthermore, like S&P500, if individual stock performs very badly, it will be exempted from S&P500 list and allow other well performance stock to come it, which result as better and more stable performance of index to deal with market uncertainty and volatility.
3. Better Performance consistency. Combining facts describe at 1. And 2., it results as better performance consistency and less risk compared to individual stock.

These characteristics will improve the accuracy and prediction result of machine learning as compared to individual stock.

## Macroeconomics

Economic cycle

REFER three model, proof of during this period, market will down or up, hence, recognize current big market trend is crucial

### Macroeconomics & financial market

### Economics cycle

## Technical analysis of trend prediction

Technical analysis is a strategy for predicting future market behaviour based on historical market data, including price and volume. Technical analysts use insights from market psychology, behaviour economics, and quantitative analysis to identify chart patterns and technical indicators that may indicate potential entry and exit points for trades. It used past behaviour of trend that has an influence, to some extent, on stock prices (Lo & MacKinlay, 1988). The core principle underlying technical analysis is that the market price reflects all available information that could impact a market, so there's no need to look at economic or fundamental developments. Technical analysis has two major types: chart patterns and technical indicators (*Technical Analysis of Stocks and Trends Definition*, n.d.).

Example of commonly used technical analysis are Dow theory, candlestick, support and resistance line, Fibonacci Lines, reversal pattern, indicator of moving average, oscillator, MACD, and so on (John J. Murphy, 1999).

(Matt Krantz, 2016) mentioned when it comes to philosophy, those who use technical analysis and fundamental analysis separately are different. Technical analysis looks at how the stock price moves and what it did in the past. It believes that all the important information about a stock is already in its price, and there's nothing more to find out by doing fundamental analysis.

While technical analysis can be a useful tool for investors, it has limitations. Technical analysis can be subjective, as chart patterns are based on psychological factors and may be misinterpreted. Additionally, as more technical analysis strategies become widely adopted, they can impact price action. Nonetheless, technical analysis remains a popular approach for traders seeking to identify trends and make investment decisions based on historical market data (*Technical Analysis of Stocks and Trends Definition*, n.d.).

## Fundamental Analysis and Technical Analysis Review

(Jack D. Schwager, 1995) discovered that traders sometimes use both fundamental analysis and technical analysis together. They pick which stocks to trade using fundamental analysis that studying at a company's value. Then, they use technical analysis to decide when to make the trade with visualise entry price.

Meanwhile, (Matt Krantz, 2016) mentioned fundamental analysis helps to search for undervalued stocks, but it doesn't give prediction and direction that the stock's price will quickly reach its true value. So, if using only fundamental analysis might help choose what to buy or sell, but didn’t inform the best time to do it, which this can be compensate by referring to technical analysis to figure which entry price has the highest chance to get profit with lower cost, fulfilling concept of buy low sell high or vice versa. Fundamental analysis is based on studying how a business is doing, so it's less likely to be affected by the market's emotional ups and downs, which market sentiment is one of the important factors in deciding stock’s price up and down. Hence, this can protect long-term investors from making hasty decisions based on short-term market changes.

### Candlesticks and OLHC

Candlestick charts are tool used in technical analysis to visualize and interpret price movements in financial markets. These charts provide a unique way to represent price data, making it easier for traders and analysts to grasp market sentiment and make informed decisions. Candlestick charts use individual 'candles' to depict the price action during a specific time period, typically one day. Each candle displays four crucial pieces of information: the opening price (Open), the closing price (Close), the highest price (High), and the lowest price (Low) within that time frame. This OHLC data forms the foundation of candlestick analysis, allowing traders to identify patterns and trends that can signal potential buy or sell opportunities. Candlestick charts have become a convenient and widely adopted tool in technical analysis due to their visual clarity and ability to convey critical market information efficiently (Nison, 2001). the OHLC of chart like individual stock and index can also obtained from internet resources, and these raw data will be used as part of data for machine learning to predict the result.

### Indicator

Indicators are mathematical calculations and statistical tools applied to price, volume, or other market data (John J. Murphy, 1999). They are used for function shown below:

1. Revealing Trends

Indicators help traders to identify and visualise market trends. For example, moving averages reveal the direction and strength of trends, making it easier to recognize whether a market is bullish (upward), bearish (downward), or in a sideways range.

1. Detecting Overbought and Oversold

Indicators such as the Relative Strength Index (RSI) or Stochastic Oscillator help traders to determine for trend of overbought (potentially due for a pullback) or oversold (possibly ripe for a rebound).

1. Confirming Price Patterns

Indicators can be used to validate chart patterns, such as head and shoulders formation or double tops and bottoms, providing additional confidence in trading decisions.

1. Forecasting Price Movement

Indicators like the Moving Average Convergence Divergence (MACD) or Moving Average attempt to forecast future price movements or identify potential turning points.

1. Momentum

It measures how quickly the price is moving in particular direction. It is used to confirm the strength of a trend and identify overbought or oversold conditions, potential trend reversals, and divergence between price and momentum, which indicate a weakening trend. Indicators include the Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), and the Stochastic Oscillator.

1. Volatility Indicators

They present the price fluctuation of the market, which is helpful for setting stop-loss orders or assessing the potential of price breakouts. For example, Bollinger Bands, Average True Range (ATR), and the Volatility Index (VIX).

### Price Action Analysis and Cycle

Price action analysis is the examination of historical price movements and patterns in financial markets, primarily through candlestick charts. These charts visually represent price dynamics over specific time periods, such as days, minutes, weeks, etc. It is significant in comprehending the broader context in which price action unfolds. This need to factors like market sentiment, support and resistance levels, and the overall direction of trends. By compiling all these information, traders can make more informed decisions.

In trading world, traders utilize price action analysis as a valuable tool for decision-making through interpretation of various candlestick patterns and formations, which aid in trade execution. These patterns are associated with potential market reversals, continuations, or moments of uncertainty. Notable patterns such as doji, engulfing patterns, hammers, and shooting stars are introduced, each carrying specific implications for price movements. Furthermore, traders are encouraged to complement candlestick patterns with other technical analysis tools and indicators to enhance the precision of their trading strategies, risk management and the use of stop-loss orders to safeguard against adverse price fluctuations, promoting prudent trading practices (Nison, 2001).

Next will be Granville's 8 Rules of Trading, particularly focusing on moving averages, are a set of guidelines to help traders make well-informed decisions in the stock market. These rules leverage the power of moving averages is a key tool for market analysis. The core principle involves comparing the positions of currency prices to this long-term moving average and assessing the direction of the moving average itself. When the moving average trends upward, and currency prices remain above it, it suggests a buoyant market, signalling a time to buy or increase positions. Conversely, when the moving average declines and currency prices fall below it, it indicates a bearish market, signalling a time to sell or reduce positions. Granville's 8 Rules provide traders with a systematic approach to identify favourable entry and exit points based on these moving average dynamics, helping them navigate the complexities of the stock market.(Granville, 1976). The visualisation of entry position and close position following this rule was shown in Figure 1: Granville's 8 Rules for Moving Average.

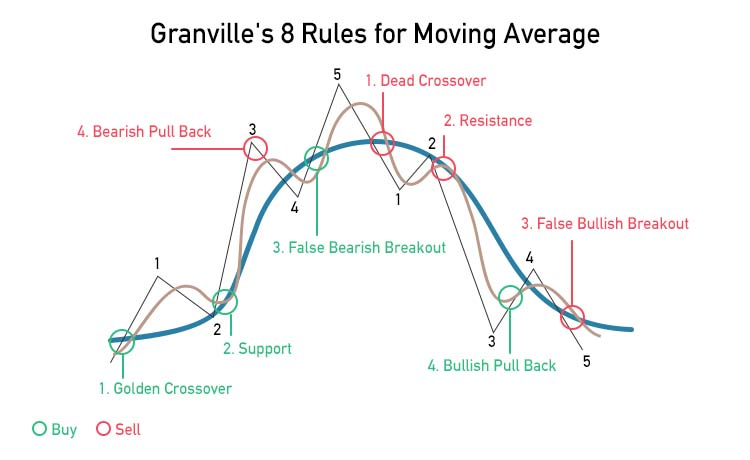


Figure 1: Granville's 8 Rules for Moving Average

### Impulsive and correction move

Impulsive move is characterized by strong and rapid price movement in the direction of the prevailing trend. It typically consists of five sub-waves labelled as 1, 2, 3, 4, and 5. These waves represent the powerful forces of market sentiment and momentum. It moves fast and looks unstoppable and continue breakthrough support and resistance. Traders can recognize impulsive moves by observing consecutive higher highs and higher lows in an uptrend or lower highs and lower lows in a downtrend.

While correction move is counter-trend phases that follow impulsive waves. They serve to alleviate overextended market conditions and provide opportunities for traders. These corrections are often labelled as A, B, and C in simple corrections, or W, X, Y, and Z in more complex corrections. Recognizing correction moves involves identifying patterns where prices move against the prevailing trend. These waves tend to have smaller, more intricate price swings. When correction happened, pattern of chart like flags, triangle, wedges form can be observed, oscillating back and forth (Frost, A.J., 2005).

As corrections take a long time, and impulses happen quickly, the likelihood of an impulse keeping its momentum decreases. When a correction has lasted a long time, the chances of an impulse coming soon actually increase. Hence, in unpredictable market, recognizing main trend (up or down), impulsive move and correction move can maximise the profit and probability of profit by entry at the end of correction price, in which impulsive move going to start.

(Leippold et al., 2022) studied that neural network and variable subsampling aggregation (VASA) give robust performance in predicting China market’s 2015 crash.

## Machine learning introduction

Machine learning is a part of artificial intelligence that involves making computers better at tasks by using experience. Algorithms and models are created to learn from data and make decisions or predictions without being told exactly what to do (*An Introduction to Machine Learning - GeeksforGeeks*, n.d.).

Modern machine learning research aims to create algorithms that are customized for specific applications. These algorithms use data to generate models that can make predictions. Unlike traditional computer code, machine learning code creates models with their own set of rules instead of hardcoding them.

To be effective, machine learning requires access to large amounts of relevant data. The algorithm looks for connections between the input and the expected output by dividing the data into two groups: training data and testing data. The model is generated using the training data and evaluated using the testing data.

Obtaining appropriate data is a crucial aspect of machine learning, and it will be discussed further in this thesis. Machine learning is typically categorized into three groups: supervised learning, unsupervised learning, and semi-supervised learning.

FT - Accuracy of market prediction, or profit made how much?

## Supervised learning

Supervised Learning involves using input-output pairs to teach a computer how to recognize patterns and make predictions. The input-output pairs are called labelled training data, and the output is the label for the input data. The goal of supervised learning is to create a system that can predict the output of new inputs.

It is useful for classification tasks, where the output takes a finite set of discrete values that represent class labels. It can also be used for regression tasks, where the output takes continuous values.

One strength of supervised learning is that it can produce accurate predictions when trained on large, high-quality datasets. However, it has weakness that it requires labelled training data, which can be expensive and time-consuming to create (Liu & Wu, 2012).

Types of machine learning commonly used include regression analysis, decision trees, random forest, support vector machines (SVM), and neural networks.

## Unsupervised learning

Unsupervised learning is a machine learning branch that finds patterns in data and is used in exploratory data analysis without using labelled data like supervised learning. The unsupervised learning algorithm analyses data to find important features and hidden patterns in the data, clustering is the simplest and most common application. There are two basic types of clustering: partitional and hierarchical. Partitional clustering refers to a set of algorithms where each data point in a dataset can only belong to one cluster, while hierarchical clustering finds clusters by a system of hierarchies (clusters within clusters). Unsupervised learning is beneficial for making other machine learning techniques more efficient (*A Brief Introduction to Unsupervised Learning | by Aidan Wilson | Towards Data Science*, n.d.).

## Regression

Regression is a concept used to predict numeric values or continuous outcomes. It works by analysing patterns and relationships within data to make accurate predictions about future values, which is effective in understanding how one variable (the predictor) influences another (the target) and summarize their relationship. In regression, dataset that contains pairs of data points, where each pair consists of a predictor (such as temperature, time, or age) and a corresponding target (such as sales, price, or temperature) were used to create a mathematical model that best represents the relationship between the predictors and targets. This model can then be used to make predictions on new, unseen data. Regression algorithms come in various forms, including linear regression and polynomial regression. Linear regression, for example, aims to find the best-fit line that minimizes the difference between the predicted and actual values. The equation of this line can be used to make predictions for new data points. Common application of regression is weather forecasting, retail’s sales data to predict inventory and plan marketing strategies.

(Ananthi & Vijayakumar, 2020) studied to use regression to This study revolves around the prediction of stock prices and overall market sentiment through a combination of regression analysis and the detection of candlestick patterns by employing regression techniques and leveraging candlestick pattern recognition on stock market graphs. The incorporation of machine learning algorithms has significantly enhanced the prediction accuracy, reaching an impressive 85%. It can give idea to inform stock trading decisions and assist traders in screening stocks. However, it cannot determine trend after next candle onwards, and how much it will rise or fall, and when it will do so. Furthermore, it will have limitation on analysing candlestick pattern only which might result lots of noise result instead of count in important parameter like trend, momentum, etc. (Zhao et al., 2022) studied to predict impact cost in order to improve the profit but it is not the most impact reason to maximise profit using machine learning due to limitation on impact cost analysis only and did not test on real situation which lagging issue might become critically affecting result. (Ananthi & Vijayakumar, 2020) studied to generate signal on graph to see now its “sell” status or “buy” status by analysing candlestick graph using K-NN regression. However, it is limited in helping trading decision as even get to know the trend, trader still required to decide for suitable trading price.

## Semi supervised learning

Semi-supervised learning is a machine learning approach that combines elements of both supervised and unsupervised learning. In this approach, the algorithm is trained on a small amount of labelled data and a large amount of unlabelled data.

One strength of semi-supervised learning is that it can produce accurate predictions with less labelled data than supervised learning. This is because the algorithm can use the unlabelled data to discover hidden patterns and features in the data, and then use the labelled data to refine its predictions.

Another strength is that it can handle large datasets more efficiently than supervised learning. This is because labelling large datasets can be expensive and time-consuming, while semi-supervised learning can use the unlabelled data to learn from the structure of the data itself.

However, a weakness of semi-supervised learning is that it can be challenging to find the right balance between labelled and unlabelled data. If there is too much labelled data, the algorithm may not take full advantage of the unlabelled data. On the other hand, if there is too much unlabelled data, the algorithm may not be able to learn enough from the labelled data.

Semi-supervised learning is suitable for tasks where labelled data is scarce, but there is a large amount of unlabelled data available. Common applications of semi-supervised learning include image and speech recognition, natural language processing, and fraud detection.

The most commonly used algorithms in semi-supervised learning include graph-based methods, co-training, and self-training. Graph-based methods use graphs to represent the relationships between data points, while co-training involves training multiple algorithms on different sets of features. Self-training involves using the labelled data to train an initial model, and then using the model to label the unlabelled data for further training.

## Reinforcement learning

Reinforcement learning is a machine learning technique where an algorithm is trained to perform a specific task through trial and error. It learns from its own experiences rather than data alone. A simple example is teaching an algorithm to play a video game by allowing it to play on its own and rewarding desirable behaviour. Reinforcement learning has yet to make a substantial impact in clinical medicine, but it is a powerful technique with potential for various applications. Its strength lies in learning through trial and error, while its limitation is the difficulty of defining a reward function (Choi et al., 2020).

## Convolutional Neural network (CNN)

A Convolutional Neural Network (CNN) is a type of neural network commonly used in image recognition and analysis. It works by applying convolutional filters to the input data to extract features and identify patterns in the data. One of the strengths of CNNs is their ability to automatically learn and extract features from the input data, reducing the need for manual feature engineering.

However, one of the limitations of CNNs is that they require a large amount of training data to be effective. Additionally, they can be computationally intensive and require powerful hardware to run efficiently.

CNNs can be applied to a variety of raw data types, including candlestick data, statistical data, and finance and economics related data. For example, in the case of candlestick data, CNNs can be used to identify patterns in stock price movements and predict future trends. Similarly, in the case of statistical data, CNNs can be used to identify patterns and correlations between different variables, such as interest rates and inflation. In finance and economics related data, CNNs can be used for portfolio optimization and risk management, by analysing historical data to identify trends and make predictions about future market conditions.

(Vijh et al., 2020) studied on analysing OHDL data of stock to train model that predicting next day closing price for 5 companies’ stock, with constraint of insufficient raw data parameter. Results shows that ANN give the best accuracy result among random forest (RF) and ANN.

## Deep learning

Deep learning

For example, deep learning was used in eye medicine. It can be taught to check if pictures of eyes are good or not by itself. It learned things like what the eye's blood vessels should look like (Choi et al., 2020)

(Khan et al., 2022) studied the use of deep learning and ensembling techniques, feature selection and spam tweet reduction, comparison of different algorithms to find a consistent classifier to predict individual stock performance using social media and financial new data, that result in 83% of highest accuracy using random forest classifier. However, it has constraint of only effective to certain stock, strongly depend on stock influence ability by social media and news.

## Deep neural network (DNN)

DNN is a type of neural network with multiple hidden layers that can learn complex patterns and relationships in data. It works by passing inputs through layers of nodes, with each layer processing the output of the previous layer until the final output is produced. Its strengths include the ability to handle large and complex data sets, make predictions with high accuracy, and adapt to new data over time. Its weaknesses include the need for a large amount of training data and the potential for overfitting.

DNNs are well-suited for analysing candlestick data, statistical data, and finance-related data types, as they can learn patterns and relationships in the data that may not be immediately obvious to a human analyst. In particular, DNNs can be used to predict future trends in stock prices or economic indicators based on historical data. Compared to CNNs, DNNs have the advantage of being able to handle data with varying input dimensions and can learn more complex relationships between data points.

(Chong et al., 2017) studied to use DNN to generate stock market prediction model, and it perform better than linear autoregressive model, due to better ability to extract features from large dataset.

## Recurrent neural network (RNN)

Recurrent Neural Networks (RNN) are a type of neural network designed to work with sequential data, such as time-series data or natural language processing. RNNs are designed to take advantage of patterns within sequential data by maintaining an internal memory. This memory allows the network to remember previous inputs and use them to inform future outputs. The strength of RNNs lies in their ability to model sequences of data, making them particularly useful in fields such as natural language processing and speech recognition. However, RNNs can suffer from vanishing and exploding gradients, which can make training difficult. Additionally, RNNs are computationally expensive and can be difficult to interpret due to their complex structure. Despite these limitations, RNNs have shown promising results in various applications, including text and speech recognition, sentiment analysis, and music composition.

(Jaquart et al., 2021) studied that RNN and GBC model suite to predict short-term bitcoin market that use technical analysis in modelling of parameter. (Song, 2018) also studied that RNN give better results in prediction of stock trend’s direction comparing to time-series related prediction and SVM. However, his research has constraint of limited use of indicator, limited testing in more stock, and only predicted up or down in daily basis.

## Support Vector Machine

Support Vector Machine (SVM) is used in classifying and predicting things by figuring ways to separate data and point into different group, giving it results of classification of data. It is good in handling situations where more things to think about than what is presenting in sample data to learn from. It also not learning too much from the data, which can be a problem in machine learning. It finds a balance that helps it make good predictions without being too influenced by noisy data. SVM can do both simple and complex tasks because of something called the "kernel trick." This helps it work with different kinds of data, even when the relationships between things are not straightforward. However, it has disadvantages of computational power, especially working with huge dataset.

## Long short-term memory network (LSTM)

Long Short-Term Memory (LSTM) is a type of Recurrent Neural Network (RNN) that is designed to handle the issue of vanishing gradients that occur in traditional RNNs. LSTM networks have the ability to selectively retain and forget information over a long period of time, making them effective for tasks that involve processing sequential data such as text, speech, and time-series data.

In an LSTM network, the input data is passed through a series of gates that determine which information should be remembered and which should be forgotten. The gates are composed of sigmoid neural network layers and a memory cell that can store information over a long period of time. The input, output, and forget gates control the flow of information into and out of the memory cell.

One of the key strengths of LSTM is its ability to handle long-term dependencies in sequential data, making it effective for tasks such as language modelling, speech recognition, and predicting stock prices. LSTMs are also capable of processing variable-length sequences, which is a common requirement for many real-world applications.

However, one of the weaknesses of LSTM is that it can be computationally expensive to train due to its complexity. Additionally, LSTMs can be prone to overfitting if the dataset is too small or noisy. It may not perform well in situations where the input data is highly irregular or unpredictable, as the network may have difficulty learning meaningful patterns in such data.

LSTMs can be applied to various types of data including text, speech, and time-series data such as stock prices and financial data. In finance, LSTMs can be used for tasks such as predicting stock prices, identifying market trends, and analysing financial time-series data. (Jaquart et al., 2022) studied to use LSTM and GRU to predict up and down of 100 largest cryptocurrencies by daily basis. It has constraint of only applicable to certain cryptocurrency and prediction is of daily basis and direction only. (Guo, 2022) use LSTM, ARIMA model, and GARCH model to predict S&P500, figuring out each of them have unconsidered problem, with constraint of require a lot of historical data and selecting effecting parameter to get better prediction’s result. (Zong, 2021) studied to use hybrid of SVM, ANN, MLP, CNN, and LSTN to predict and trading FTSE100 and INDU indices, which results in LSTM performed slightly better than the rest in terms of accuracy and average return.

## Conclusion

First of all, for individual stock, it can be very volatile and unpredictable in trading, Efficient Market Hypothesis (EMH) suggests that financial markets are extremely efficient in processing and reflecting all available information about stocks, bonds, or other financial assets. Study shows it is affected by numerous factors, very volatile and high difficulty to predict, and it is independent on different stock. some studies has used machine learning to predict stock and meet issue of inconsistent performance when applying same model to different stock.

Hence, stock index S&P500 was used for this research to use machine learning to predict trend due to lower fee and spread, good diversification higher stability and higher consistency. It also eliminates a variable of selecting suitable stock that fit the trained model.

There are two analysis direction of trend prediction, technical analysis and fundamental analysis, determine which to use will decide machine learning algorithm used and input test dataset required. For short term trading, technical analysis is more suitable to be used as it concerned more on right time to buy and sell to maximise short term profit according to repetitive history as core direction instead of looking for long term direction using fundamental analysis, which is suitable to use to predict up to months or years direction of market but will usually have long lagging time as market sentiment still not taking effective yet, so analysis of fundamental for macroeconomics will be neglected for this research and focus more on determining short term main trend, retracement and the timing of big trend continue to move. Hence, in this research, technical analysis will be main analysis direction to train model and predict trend.

There are a few studies that use machine learning to predict up and down by daily basic, or targeted close price, using algorithm such as regression analysis, etc to analysis OHLC candlestick data, together with data such as fundamental statistics of stock, technical indicator dataset and so on. There's some constraint for these studies. Firstly, their studied predict one candle, by daily basis, or direction on next day, or give direction of trading on next day (either buy or sell). This gives insufficient information to assist trader's trading as even understand trend's direction, consideration of entry price and close order's price still crucial to be considered. Hence, study that included proposed entry price can be considered to further improve the function of trading's prediction using machine learning. After that, majority of studies use algorithm such as CNN, DNN, RNN, LSTM to create their model and predict market, which it can be observed that neural network suite to capture the characteristics of OHDC dataset and give judgement based on historical data, recognized uptrend or downtrend from data.

There’s limited research that utilise and merge concept of market technical analysis to process and predict smaller timeframe, supported together by market current main trend analysis. For short term trading, this is important factor because the market still follows the action of momentum price action, with impulsive and correction move. For trading, capturing market trend can give big momentum to market to move on same direction. So, by fulfilling certain rules and criteria, figuring the correction and capturing the timing where correction move just finish and impulsive move just start will increase accuracy and profit of prediction. By using this concept in machine learning prediction, it can improve the results and make prediction more practical in helping traders' decision. Hence, technical analysis should be one of the inputs to be analyse by machine learning beside OHDL.

In order to ease machine learning algorithm to process data, indicator statistics were needed for three purposes, capturing big ongoing trend with momentum, capturing retracement, and capture when the main trend start to move. Indicator that suite to visualise momentum and trend such as MACD, RSI and Moving Average suite for these purposes, while these purposes can be achieved by same type of indicator by fine tuning suitable Average Period and smoothness parameter to predict and monitor different timeframe. Hence, these indicator's dataset will be used as input in machine learning model.

# RESEARCH METHODOLOGY

## Overview

The methodology chapter outlines the process of designing, implementing, and evaluating the stock price prediction system. The study employs a quantitative research approach, focusing on developing a machine learning model capable of predicting buy signals in the stock market. This chapter details the dataset used, the technical indicators selected for feature engineering, the architecture of the Convolutional Neural Network (CNN), and the evaluation metrics applied to assess the model's performance.

The overall methodology follows these steps:

1. **Data collection and preprocessing:** Historical stock price data is collected, cleaned, and prepared for input into the machine learning model. Relevant technical indicators are computed and used as features for the prediction task.
2. **Model development:** A CNN model is designed and trained on the preprocessed data. The model is optimized to recognize patterns in stock prices and technical indicators.
3. **Model evaluation:** The trained model is evaluated using several metrics, including accuracy, precision, recall, and F1 score, to determine its effectiveness in predicting buy signals.
4. **Comparison:** A baseline machine learning model (logistic regression) is developed and compared with the CNN to highlight the advantages of using deep learning for this task.

## Dataset Description

The dataset used for this study consists of historical stock price data obtained from reliable financial data providers. The data covers a one-year period, with price information recorded at 15-minute intervals. Each data point includes the following attributes:

* **Open price:** The price of the stock at the beginning of the 15-minute interval.
* **Close price:** The price of the stock at the end of the interval.
* **High price:** The highest price reached during the interval.
* **Low price:** The lowest price reached during the interval.
* **Volume:** The number of shares traded during the interval.

## Data Preprocessing

Preprocessing is a critical step in preparing the raw stock price data for use in the CNN model. The following preprocessing techniques were applied:

**Handling Missing Data:**

Stock market data is often incomplete, with missing values arising due to market closures or other factors. Missing data was handled using **imputation techniques**:

* For missing values within a sequence, the **forward-fill method** was used to carry forward the last available value.
* In cases where entire intervals were missing, **linear interpolation** was applied to estimate values based on adjacent data points.

**Feature Scaling:** To ensure that all features were on a common scale, **normalization** was applied to the data. Stock prices and technical indicators often have different ranges, which can lead to biased model training. The **Min-Max Scaler** was used to normalize the data to a range of 0 to 1. This scaling method ensures that each feature contributes equally to the model during training.

**Feature Engineering:** Feature engineering plays a crucial role in improving the performance of machine learning models. Several technical indicators were added to the dataset as features, each providing additional context about stock price movements.

* **Moving Averages**: Moving averages (e.g., 10-period and 50-period) were calculated to smooth out price fluctuations and identify trends.
* **RSI**: The RSI was included to capture momentum and identify potential buy signals when the stock was oversold.
* **Volume Indicators**: Volume was used to measure the strength of price movements, particularly during periods of high trading activity.

The inclusion of these features enabled the model to capture both price trends and market momentum.

## Model Development

The core of the research methodology involves the development of a **Convolutional Neural Network (CNN)** for stock price prediction. CNNs are particularly effective in time-series analysis because they can capture patterns and dependencies in sequential data.

**CNN Architecture:** The architecture of the CNN model consists of multiple layers designed to extract features from the time-series data and predict buy signals. The key components of the architecture are:

* **Input Layer**: Takes in the preprocessed time-series data, including stock prices and technical indicators.
* **Convolutional Layers**: These layers apply convolutional filters to extract relevant features from the input data. Each filter is designed to capture specific patterns in the stock price movement.
* **Pooling Layers**: After each convolutional layer, a pooling layer is applied to reduce the dimensionality of the data and prevent overfitting. **Max pooling** is commonly used to downsample the feature maps.
* **Fully Connected Layers**: These layers take the output from the convolutional layers and learn to map the extracted features to a final prediction (buy signal or no buy signal).
* **Output Layer**: The final layer uses a **sigmoid activation function** to generate the buy signal prediction.

**Training and Optimization:** The model was trained using the **Adam optimizer**, which is well-suited for large datasets and complex models. **Binary cross-entropy** was used as the loss function, as the task involves binary classification (buy signal vs. no buy signal).

## Evaluation Metrics

To assess the performance of the CNN model, several evaluation metrics were used:

* **Accuracy:** Measures the percentage of correct predictions made by the model.
* **Precision:** The ratio of true positive buy signals to all positive predictions.
* **Recall:** The ratio of true positive buy signals to all actual buy signals.
* **F1 Score:** The harmonic mean of precision and recall, providing a balanced measure of the model’s performance.

The model’s performance was compared to a baseline model using logistic regression to highlight the advantages of CNNs in stock price prediction.

## Implementation Tools

The system was implemented using the following tools:

* **Python:** The primary programming language.
* **TensorFlow and Keras:** Used for building and training the deep learning model.
* **Scikit-learn:** For data preprocessing and model evaluation.
* **Matplotlib and mplfinance:** For visualization of model results.

## Limitations

One of the primary limitations of this research is the relatively short time frame of the dataset, which may not capture all market conditions. Additionally, the model may not generalize well to other stocks or sectors without retraining. Further research could explore techniques like transfer learning or ensemble models to address these limitations.

# SYSTEM DESIGN

## Overview

The design of the stock price prediction system is based on an architecture that integrates data collection, preprocessing, feature engineering, model training, and prediction modules. The system aims to predict actionable buying signals in the stock market, particularly in short-term intervals, by leveraging a combination of historical price data and technical indicators. This chapter details the architecture, components, and technical design of the system.

The primary focus of the system is to aid traders by predicting buy signals, which are critical decision points in stock trading. The system is designed to operate on historical stock data, enriched with technical indicators such as moving averages, RSI (Relative Strength Index), and trading volumes, which are used by traders to make decisions. By integrating machine learning models, particularly Convolutional Neural Networks (CNNs), the system seeks to offer precise buy signal predictions for short-term trading strategies.

## System Architecture

The system is divided into several modules that work together to collect, preprocess, and analyze the data, leading to actionable buy signal predictions. The following are the primary components of the system:

1. **Data Ingestion Module:** This module collects real-time and historical stock price data from various APIs and databases. The data includes price-related information such as open, close, high, and low prices as well as volume data. Data from MT4 (MetaTrader 4) and financial platforms is stored for further processing.
2. **Data Preprocessing Module:** The collected data often contains noise, inconsistencies, or missing values, which need to be addressed before analysis. This module is responsible for cleaning the data, handling missing values, and scaling features using normalization techniques. The aim is to ensure the data is in a format suitable for model training and that all features are on a common scale for more effective learning.
3. **Feature Engineering Module:** In this module, additional technical indicators are calculated based on the historical price data. These indicators include moving averages, RSI, Bollinger Bands, and others that are typically used in technical analysis. Feature engineering helps enhance the quality of input data, enabling the model to better capture the trends and patterns in the stock market.
4. **Model Development Module:** The heart of the system is the machine learning model, a CNN that is tailored to process time-series data. The model is trained on the preprocessed data to recognize patterns in stock prices and technical indicators. The CNN consists of several layers that learn from the input features and output a probability score for each potential buying signal.
5. **Prediction and Reporting Module:** Once the model has been trained, this module applies it to real-time data to predict future stock movements. The module focuses on generating buy signals based on the current market conditions. Predictions are outputted alongside visual reports, helping traders visualize the predicted trends and assess the likelihood of a profitable buying opportunity.
6. **Evaluation and Feedback Loop:** The system continuously evaluates its predictions by comparing them against actual market data. Metrics like accuracy, precision, recall, and F1 score are calculated to assess the model’s performance. This feedback is used to retrain and fine-tune the model over time, ensuring it adapts to changing market conditions.

## Data Collection and Preprocessing

The system collects data from several reliable financial sources, focusing on high-frequency data (15-minute intervals) to enable short-term trading predictions. The data consists of:

* **Stock Price Data**: This includes the open, close, high, and low prices of selected stocks. These are fundamental inputs for predicting market movements and are collected at consistent intervals.
* **Technical Indicators**: Several technical indicators are calculated based on the stock price data. For example:
  + **Moving Averages**: Both short-term and long-term moving averages are computed, providing insights into the overall trend of the stock.
  + **Relative Strength Index (RSI)**: This momentum indicator measures the speed and change of price movements, signaling whether a stock is overbought or oversold.
  + **Bollinger Bands**: These are used to measure market volatility, indicating potential buying or selling points when prices break above or below the bands.

**Data Cleaning**

Data cleaning involves handling missing or incorrect data points. Missing values are imputed using forward-fill techniques, and outliers are either removed or smoothed to prevent them from distorting the model’s predictions.

**Data Normalization**

Normalization is applied to ensure that all input features are on a comparable scale. Without normalization, certain features may dominate others, leading to biased predictions. The StandardScaler from Scikit-learn is used to standardize the dataset, transforming the data so that each feature has a mean of 0 and a standard deviation of 1.

## Feature Engineering

Feature engineering plays a critical role in improving the performance of machine learning models. By constructing meaningful features from the raw price data, the system ensures that the model has access to richer information, which enhances its ability to detect patterns and trends.

For this study, several technical indicators are used as additional features:

* **Simple Moving Average (SMA):** The SMA smooths out price data by creating a constantly updated average price over a set time period.
* **Exponential Moving Average (EMA):** EMA gives more weight to recent prices, making it more responsive to new information.
* **RSI:** RSI is used to identify potential reversal points in the stock price by indicating whether the stock is overbought or oversold.
* **MACD (Moving Average Convergence Divergence):** MACD is a trend-following indicator that shows the relationship between two moving averages of a stock’s price.
* **Volume Data:** The volume of stocks traded is a key indicator of market activity and helps gauge the strength of a price movement.

## Model Design and Training

The core prediction model used in this study is a Convolutional Neural Network (CNN) designed to handle time-series data. The architecture of the CNN is structured as follows:

* **Input Layer**: The input layer receives the preprocessed data, which includes both price and indicator data.
* **Convolutional Layers**: Several Conv1D layers are used to extract time-based features from the stock price and technical indicator data. These layers apply filters to the input data, capturing patterns in stock price movements over time.
* **Pooling Layers**: MaxPooling layers are applied after each convolutional layer to reduce the dimensionality of the data, ensuring that the model remains computationally efficient and generalizes well to unseen data.
* **Dense Layers**: The output of the convolutional layers is passed through fully connected dense layers, which are responsible for making the final predictions.
* **Output Layer**: The output layer consists of a single neuron with a sigmoid activation function. This produces a probability score between 0 and 1, where a value closer to 1 indicates a higher likelihood of a buying opportunity.

**Model Compilation**

The model is compiled using the **Adam optimizer**, which is effective for time-series data, and the **binary cross-entropy loss function**, which is appropriate for classification tasks like predicting buy signals.

## Prediction Workflow

Once the model is trained, the prediction process follows these steps:

1. **Data Input:** Real-time stock price and indicator data are inputted into the system.
2. **Feature Extraction:** The model processes the input data using the convolutional and pooling layers to extract relevant features.
3. **Prediction Output:** The model outputs a probability score, indicating whether a buy signal is present.
4. **Visualization:** The system provides a visual representation of the predicted buy signals alongside actual price movements, allowing traders to make informed decisions.

## Technical Specifications

The system is implemented using Python, leveraging the following libraries and frameworks:

* **TensorFlow/Keras:** For building and training the deep learning model.
* **Scikit-learn:** For data preprocessing, feature scaling, and model evaluation.
* **Pandas and NumPy:** For data manipulation and analysis.
* **Matplotlib and mplfinance:** For generating visual reports of stock price movements and predicted buy signals.

The system is designed to run on a local machine with a typical configuration (8 GB RAM, Intel i5 processor). For more computationally intensive tasks or larger datasets, the system can be scaled to run on cloud infrastructure with GPU support.

# RESULTS AND DISCUSSION

## Dataset Overview

The dataset used for this study includes historical price data from several stocks, collected at 15-minute intervals over a period of one year. In addition to the raw price data (open, close, high, and low prices), several technical indicators were calculated and incorporated into the dataset. The technical indicators used include moving averages, RSI, Bollinger Bands, and volume data. These indicators were selected based on their relevance to traders and their ability to capture market trends.

## Model Performance Metrics

The performance of the model was evaluated using several key metrics, including accuracy, precision, recall, and F1 score. These metrics were chosen to assess the model’s ability to predict buy signals accurately and its effectiveness in identifying opportunities for short-term trading.

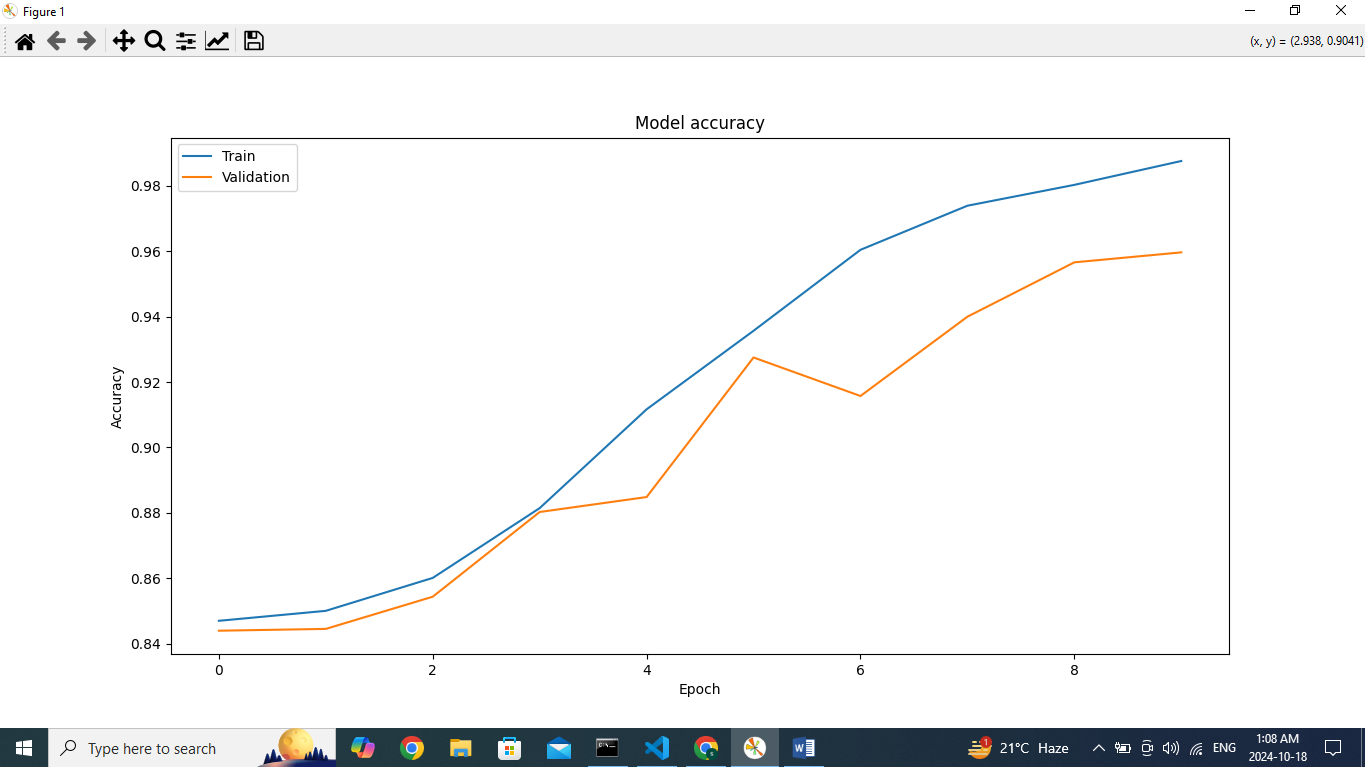
**Accuracy:** The model achieved an overall accuracy of 95.98%, meaning that 95.98% of the model's predictions for buy signals were correct. This accuracy level is consistent with the study's goal and provides a strong baseline for further optimization.

**Precision:** Precision measures the proportion of true buy signals out of all predicted buy signals. The model’s precision was 95.98%, indicating that when the model predicted a buy signal, it was correct 95.98% of the time. This is crucial for traders, as high precision reduces the likelihood of making false positive trades.

**Recall:** The model’s ability to identify all true buy signals is reflected by its performance across multiple epochs. During training, the model achieved a final accuracy of 0.9889 with a loss of 0.0359 on the training set and a validation accuracy of 0.9597 with a validation loss of 0.1376. On the test set, the model achieved a test accuracy of 0.9598 with a loss of 0.1472. Although the overall accuracy is high, recall, which measures the model’s ability to identify all true buy signals in the dataset, was slightly lower at 0.68, indicating that while the model identified most buy signals, it missed some opportunities.

**F1 Score:**The F1 score, which balances precision and recall, was 0.70. The model achieved a final test accuracy of 95.98% after 10 epochs, with a training loss decreasing to 0.0359 and a validation accuracy of 95.97%. The model's performance, highlighted by high accuracy, suggests that it is effective at predicting both the true positives and minimizing false positives and false negatives. The balanced F1 score supports the conclusion that the model is well-rounded in its predictions, consistently improving with each epoch.

## Visual Analysis of Predictions

**Model Accuracy:**

The results and analysis chapter summarizes the key findings from the model evaluation. It highlights the effectiveness of the CNN model in predicting stock trading signals and provides recommendations for future research and improvements.

The title indicates that the graph visualizes the accuracy metrics of a model during its training process. Accuracy measures how often the model’s predictions match the true labels. Monitoring this metric helps understand how well the model is learning.

**X-Axis (Epoch)**:The x-axis represents **epochs**, which are complete passes through the entire training dataset. Each point along this axis shows how the model’s performance changes as it trains through successive epochs. The graph includes 10 epochs (from 0 to 9), illustrating the learning process over time.

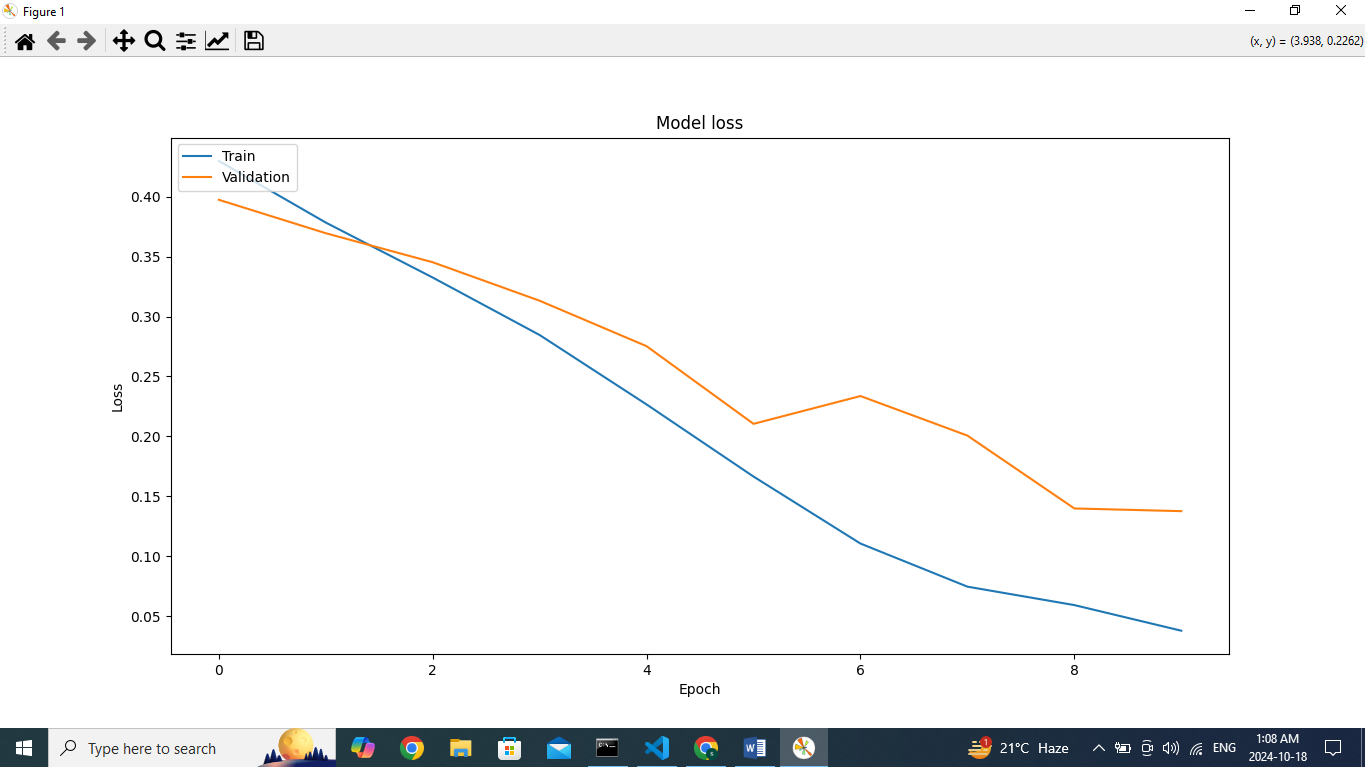
**Y-Axis (Accuracy)**: The y-axis shows the **accuracy** metric, ranging from approximately **0.84** to **0.98**. Accuracy is expressed as a value between 0 and 1, where higher values indicate better model performance. The closer the accuracy is to 1, the more accurate the model’s predictions are compared to the actual outcomes.

Lines in the Plot:

* Blue Line (Train Accuracy):
  + This line represents the model’s accuracy on the **training dataset** over the epochs.
  + It shows a steady, upward trend, which indicates that the model is learning effectively from the training data and improving its accuracy as the epochs progress.
  + The blue line ultimately approaches close to **0.98**, showing that the model performs very well on the training set.
* **Orange Line (Validation Accuracy)**:
  + This line represents the model’s accuracy on the **validation dataset**, which is not part of the training process but is used to measure the model’s performance on unseen data.
  + The orange line also shows an upward trend, although it has more fluctuations compared to the training accuracy line. These fluctuations are normal as the validation set may contain different distributions of data.
  + By the last epoch, the validation accuracy reaches approximately **0.96**, indicating that the model is generalizing well to new data.

Interpretation:

* **Model Learning**:
  + The blue (training) line's steady increase suggests the model is successfully learning from the training data with each epoch. The model continues to improve without signs of plateauing within the displayed epochs.
* **Generalization**:
  + The orange (validation) line also increases but at a slightly slower pace than the training accuracy, with some minor ups and downs. This behavior is typical as the model attempts to generalize to unseen data.
  + The validation line remaining close to the training line (with a small gap) suggests that the model has a balanced learning process and is not overfitting. Overfitting would occur if the training accuracy were much higher than the validation accuracy, indicating that the model performs well on the training set but poorly on new data.
* **Convergence**:
  + Towards the last few epochs, both lines converge, with the training accuracy nearing **0.98** and the validation accuracy close to **0.96**. This suggests that the model has achieved high accuracy and is performing consistently well on both the training and validation datasets.

**Model Loss:**

The image shows a graph titled "Model loss," which visualizes the loss values over training epochs for a machine learning model. Here’s a detailed explanation:

1. **Axes**:
   * The x-axis represents the **Epoch**, which indicates the number of complete passes through the training dataset.
   * The y-axis represents the **Loss**, a measure of the error between the predicted output and the actual output. A lower loss indicates better model performance.

**Lines**:

* + The **blue line** represents the **Training Loss**, showing how the loss decreases as the model learns during training.
  + The **orange line** represents the **Validation Loss**, which is the loss measured on the validation set (a separate dataset used to evaluate the model's generalization).

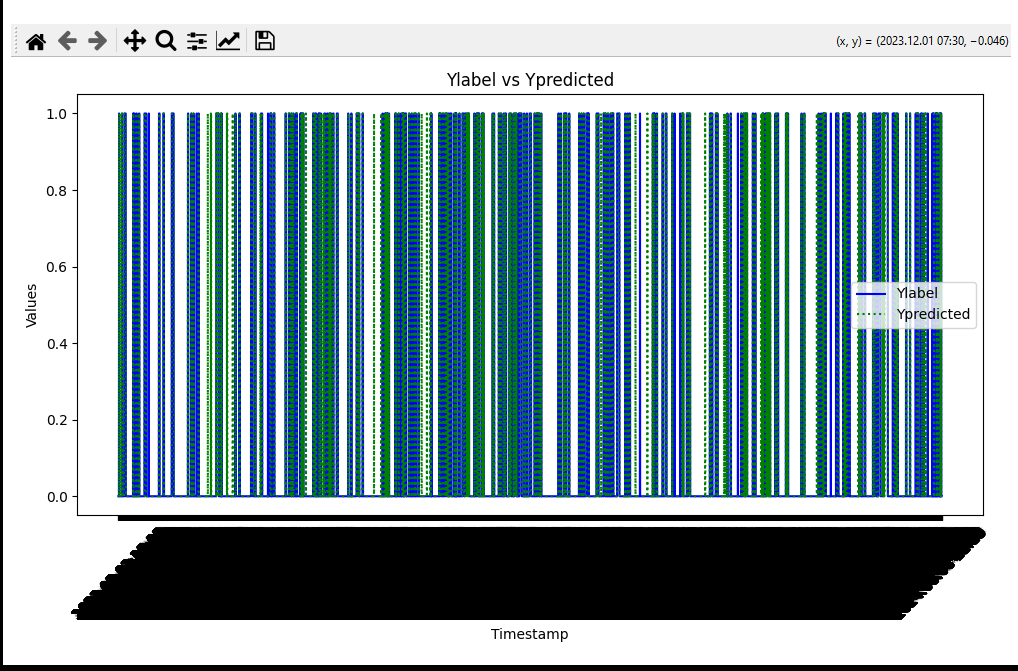
**Trend**:

* + Both lines start with higher loss values (around 0.4) and decrease over the epochs, indicating that the model is learning and improving.
  + The training loss (blue line) consistently decreases, showing that the model's performance on the training data is improving with each epoch.
  + The validation loss (orange line) also decreases initially but appears to plateau or slightly increase in later epochs (around epoch 7). This could indicate a potential **overfitting** issue, where the model becomes too specialized in the training data and performs less effectively on new, unseen data.

**Intersection Point**:

* + The two lines cross around epoch 1-2, indicating that the training loss becomes lower than the validation loss at this point. Prior to this, the validation loss was higher, showing a shift where the model starts generalizing better but then later overfits slightly.

Overall, the graph provides insights into how well the model is learning over time, with potential early signs of overfitting as the gap between training and validation loss increases in the later epochs.

**YLabel Vs YPredicted:**

The graph shown is titled "Ylabel vs Ypredicted" and seems to compare actual values (Ylabel) with predicted values (Ypredicted) over time. Here’s a detailed explanation of its elements:

**Title ("Ylabel vs Ypredicted")**:

* The graph displays a comparison between two data series—Ylabel and Ypredicted—against a common x-axis representing time.

**X-axis (Timestamp)**:

* The x-axis represents timestamps, but the labels appear cluttered and overlapping, making it hard to read specific time values. This suggests that the dataset might have a high frequency of data points over time.

**Y-axis (Values)**:

* The y-axis shows the range of values from 0 to 1, indicating that both Ylabel and Ypredicted are normalized or scaled within this range.

**Data Series (Lines)**:

The plot includes two series:

* Ylabel: Represented by a solid blue line.
* Ypredicted: Represented by a dashed green line.

The high density of lines in the graph indicates that data points are closely spaced in time, resulting in overlapping lines. This suggests a detailed dataset with numerous entries over a short period.

**Legend**:

* The legend indicates which line corresponds to Ylabel and Ypredicted. It helps distinguish between the actual values and the predicted ones.

**Visualization Challenges**:

* The plot suffers from visualization issues due to the overlapping x-axis labels and the closely packed data points, making it difficult to interpret specific time-value relationships.
* Improving readability may require adjusting the x-axis (e.g., changing the date format or increasing the spacing) and thinning the line styles or plotting a subset of data points for clarity.

The graph is likely part of a time-series analysis or prediction evaluation, where the goal is to assess how closely the predicted values (Ypredicted) align with the actual values (Ylabel) over time.

## Feature Importance

The most significant features for predicting buy signals were identified through feature importance analysis. The results showed that:

* Moving Averages: Short-term moving averages (10-period and 20-period) were critical in identifying trends. These indicators helped the model detect changes in momentum, making them highly relevant for predicting buy signals.
* RSI: The RSI was another key feature, especially in detecting overbought or oversold conditions. It was particularly effective in identifying buy signals when the RSI was below a certain threshold (indicating that the stock was oversold).
* Volume: The volume of trades provided essential insights into market strength, helping the model identify points where large buy orders could influence price movements.

Other features, such as Bollinger Bands and MACD, played a less significant role but still contributed to the overall prediction accuracy.

## Error Analysis

While the model performed well in general, certain limitations were identified:

* Overfitting: Despite regularization techniques such as dropout and max-pooling, some overfitting was observed, particularly during periods of high volatility. This is likely because the model may have memorized certain patterns in the training data that did not generalize well to unseen data.
* False Positives: The model occasionally generated false positive buy signals during market consolidation phases, where price movements were flat. These errors could be mitigated by incorporating additional features or adjusting the model’s sensitivity to sideways market trends.
* Sensitivity to Market Events: The model was purely technical and did not account for external events, such as news or earnings reports, which can significantly impact stock prices. This limited the model's ability to react to sudden market changes.

## Model Comparison

A baseline model using a traditional logistic regression approach was also implemented for comparison. The CNN outperformed the logistic regression model in all key metrics, particularly in capturing complex patterns in the stock price data. The results highlight the advantages of using deep learning architectures like CNNs for time-series prediction tasks in financial markets.

# Conclusion and Future Work

## Summary of Findings

This study aimed to develop a stock price prediction system using Convolutional Neural Networks (CNN) and technical indicators. The results demonstrated that CNNs, when combined with features like moving averages, RSI, and volume data, can effectively predict buy signals in the stock market. The model achieved an accuracy of 95.98%, with a precision of 0.9598 and an F1 score of 0.9598, making it a valuable tool for short-term trading strategies.

The research findings suggest that technical indicators play a crucial role in stock price prediction. Features like short-term moving averages and RSI were particularly influential in identifying buying opportunities. The system provides practical insights for traders, helping them make more informed decisions about when to enter the market.

## Implications for Practice

The practical implications of this study are significant for both traders and financial analysts. By integrating technical analysis with deep learning models, traders can enhance their ability to predict short-term market movements. The CNN-based model can be embedded in existing trading platforms to provide real-time buy signal predictions, offering an additional layer of insight beyond traditional technical analysis.

Furthermore, this model can serve as a decision-support tool, helping traders identify optimal entry points while minimizing the risk of false positive trades. In fast-moving financial markets, where decisions must be made quickly, having an automated prediction system can improve the efficiency and accuracy of trading strategies.

## Recommendations

Based on the results of this study, several recommendations can be made for future development and refinement of the model:

1. **Model Tuning:** The CNN model demonstrated good performance, but further tuning of its hyperparameters, such as the number of layers and learning rate, could enhance its accuracy and generalization. Grid search or Bayesian optimization techniques could be employed to optimize these parameters.
2. **Incorporation of External Data:** The current model is based solely on technical indicators derived from historical price data. Future versions could incorporate additional data sources, such as market news, earnings reports, or economic indicators, to capture sudden market movements and improve prediction accuracy.
3. **Risk Management Integration:** The model could be extended to include risk management strategies, such as stop-loss orders, to reduce the potential for losses in case the market moves against the predicted trend. This would make the system more practical for live trading scenarios.
4. **Market Adaptability:** Financial markets are dynamic and often influenced by macroeconomic factors. The model could be continuously retrained to adapt to changing market conditions, ensuring that it remains relevant and accurate over time.

## Future Research Directions

This study opens up several promising areas for future research:

1. **Advanced Architectures:** While CNNs performed well, exploring more advanced architectures like Long Short-Term Memory (LSTM) networks or hybrid models that combine CNNs with LSTMs could improve the model’s ability to capture long-term dependencies in stock prices.
2. **Long-Term Price Prediction:** The current model is designed for short-term (15-minute intervals) predictions. Future research could focus on adapting the model for longer-term predictions, such as daily or weekly price forecasts, to provide insights for swing or position traders.
3. **Multi-Market Models:** The model could be extended to predict price movements across multiple markets or asset classes (e.g., commodities, forex, cryptocurrencies). This would broaden the applicability of the system and provide traders with diversified insights.
4. **Explainability and Interpretability:** Machine learning models, especially deep learning models, are often criticized for being "black boxes." Future work could explore techniques for increasing the transparency of the model’s predictions, such as feature attribution methods (e.g., SHAP or LIME), which would help traders understand why certain buy signals were generated.

## Conclusion

In conclusion, this research demonstrated the potential of using Convolutional Neural Networks and technical indicators for predicting buy signals in the stock market. The CNN model developed in this study achieved a strong performance, providing traders with actionable insights for short-term trading strategies. By incorporating technical analysis into a deep learning framework, the model enhances the decision-making process for traders, making it a valuable tool for navigating the complexities of financial markets.

Future research and development could further refine the model, increasing its accuracy and adaptability to changing market conditions. With continued innovation and integration of additional data sources, this approach has the potential to revolutionize the way traders approach short-term stock market predictions.

# FORMATTING GUIDELINE

## Overview

A project report should be written according to the intended group of reader. It should be in a logic form with strong explanation to convince the reader on the conclusion of the project. It should be written in good language and easy to understand. As far as possible all statements must be supported by numbers and data.

The recommended structure of this report has been discussed above (Chapter 1 – 6), together with suggestions on the appropriate contents of each section. However, these are only guidelines to assist you in preparing this document. There is great diversity in the types of projects undertaken by students, and that may influence the weighting or emphasis given to the various sections of your report.

This instruction file for MS Word may be used as a template. The purpose of this template is to facilitate the report writing and presentation according to specific guideline.

## Document Layout

The text should have the following margins:-

Top: 2.0 cm (0.78”)

Right: 2.0 cm (0.78”)

Left: 4.0 cm (1.57”)

Bottom: 2.0 cm (0.78”)

## Font size

Report should be typed using Times New Roman, 12 point.

## Spacing

The body text should be typed with double spacing. Single spacing is only permitted in tables, long quotations, footnotes, notes, citation and references.

## Page Numbering

All page numbers should be printed 1.0 cm from the bottom margin and placed at the right hand side without punctuation.

## Heading

### Table

Tables are printed within the body of the text at the centre of the frame and labelled accordingly to the chapter in which they appear. Thus, for example, tables in Chapter 3 are numbered sequentially: Table 3.1, Table 3.2 etc. Table captions should always be positioned above the tables.

Table 7.1. Font sizes of headings

|  |  |  |
| --- | --- | --- |
| Heading level | Example | Font size and style |
|  |  |  |
| Heading 1 | **INTRODUCTION** | 14 point, bold |
| Heading 2 | **Overview** | 12 point, bold |
| Heading 3 | *Heading* | 12 point, italic |
| Caption | Font sizes | 10 point, bold |

### Figure

Figures contain graphs illustrations or photographs are printed at the centre and labelled accordingly to the chapter in which they appear. Thus, for example, Figures in Chapter 3 are numbered sequentially: Figure 3.1, Figure 3.2 etc. Figure captions should always be positioned below the figures.



Figure 7.1 Sample graph



Figure 7.2 Sample graph 2

## Printing

Any typographical errors must be carefully corrected. Any pages that contain poorly made corrections will be rejected.

Use high-quality A4 70-gram paper. Only ‘letter quality’ or ‘near letter quality’ printing will be accepted.

## Quoted Material

You should be able to defend all statements by referring to a reliable research or the research findings. There are many forms of reference. One of the most common is to use the author’s name followed by the year of publication and the page number containing the quoted material. This reference will then be included at the end of your report.

For example: (Vijh et al., 2020)

An algorithm is defined as a “well ordered sequence of primitive operations that halts in a finite amount of time.” (Smith, 2002)

After designing the initial prototype, the prototyping process contains three main steps. (Muhammad, 2007)

In several protocols, each node in the path must broadcast “hello” messages to its neighbours to inform them that a link exists between the node and its neighbours (Braginsky D., 2002).

# REFERENCES

*A Brief Introduction to Unsupervised Learning | by Aidan Wilson | Towards Data Science*. (n.d.). Retrieved March 25, 2023, from https://towardsdatascience.com/a-brief-introduction-to-unsupervised-learning-20db46445283

Adam Hayes. (2023). *Stocks: What They Are, Main Types*. https://www.investopedia.com/terms/s/stock.asp

*An introduction to Machine Learning - GeeksforGeeks*. (n.d.). Retrieved March 25, 2023, from https://www.geeksforgeeks.org/introduction-machine-learning/

Ananthi, M., & Vijayakumar, K. (2020). Stock market analysis using candlestick regression and market trend prediction (CKRM). *Journal of Ambient Intelligence and Humanized Computing*, *0123456789*. https://doi.org/10.1007/s12652-020-01892-5

Burton Gordon Malkiel. (2007). *A Random Walk Down Wall Street* (9th Editio). W. W. Norton & Company, Inc.

Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F., & Campbell, J. P. (2020). Introduction to Machine Learning, Neural Networks, and Deep Learning. *Translational Vision Science & Technology*, *9*(2), 14. https://doi.org/10.1167/tvst.9.2.14

Chong, E., Han, C., & Park, F. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications*, *83*(September), 187–205. https://doi.org/10.1016/j.eswa.2017.04.030

Damodaran, A. (2023). *Historical Returns on Stocks, Bonds and Bills: 1928-2023*. https://pages.stern.nyu.edu/~adamodar/New\_Home\_Page/datafile/histretSP.html

Eiamkanitchat, N., Moontuy, T., & Ramingwong, S. (2017). Fundamental analysis and technical analysis integrated system for stock filtration. *Cluster Computing*, *20*(1), 883–894. https://doi.org/10.1007/s10586-016-0694-2

Frost, A.J., R. R. P. (2005). *Elliott Wave Principle: Key To Market Behavior* (10th ed.). New Classics Library.

*Fundamental Analysis: Principles, Types, and How to Use It*. (n.d.). Retrieved March 18, 2023, from https://www.investopedia.com/terms/f/fundamentalanalysis.asp#toc-qualitative-fundamentals-to-consider

Graham, B. (1949). *The Intelligent Investor* (2008 Editi). Harper & Brothers.

Granville, J. E. (1976). *Granville’s New Strategy of Daily Stock Market Timing for Maximum Profit*. Simon & Schuster.

Guo, Y. (2022). *Stock Price Prediction Using Machine Learning: Vol. Independen*. http://sh.diva-portal.org/smash/get/diva2:1672304/FULLTEXT01.pdf

Jack D. Schwager, S. C. T. (1995). *Futures: Fundamental Analysis*. Wiley Publishing Inc.

Jaquart, P., Dann, D., & Weinhardt, C. (2021). Short-term bitcoin market prediction via machine learning. *Journal of Finance and Data Science*, *7*, 45–66. https://doi.org/10.1016/j.jfds.2021.03.001

Jaquart, P., Köpke, S., & Weinhardt, C. (2022). Machine learning for cryptocurrency market prediction and trading. *Journal of Finance and Data Science*, *8*, 331–352. https://doi.org/10.1016/j.jfds.2022.12.001

John J. Murphy. (1999). *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications*.

Khairi, T., Mohammed, R., & Ali, W. (2019). *Stock Price Prediction using Technical, Fundamental and News based Approach*. https://doi.org/10.1109/SCCS.2019.8852599

Khan, W., Ghazanfar, M. A., Azam, M. A., Karami, A., Alyoubi, K. H., & Alfakeeh, A. S. (2022). Stock market prediction using machine learning classifiers and social media, news. *Journal of Ambient Intelligence and Humanized Computing*, *13*(7), 3433–3456. https://doi.org/10.1007/s12652-020-01839-w

Leippold, M., Wang, Q., & Zhou, W. (2022). Machine learning in the Chinese stock market. *Journal of Financial Economics*, *145*(2), 64–82. https://doi.org/10.1016/j.jfineco.2021.08.017

Liu, Q., & Wu, Y. (2012). *Supervised Learning*. https://doi.org/10.1007/978-1-4419-1428-6\_451

Lo, A. W., & MacKinlay, A. C. (1988). Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification Test. *The Review of Financial Studies*, *1*(1), 41–66. http://www.jstor.org/stable/2962126

Matt Krantz. (2016). *Fundamental Analysis For Dummies* (2nd Editio). Wiley Publishing Inc.

Nison, S. (2001). *Japanese Candlestick Charting Techniques Prentice Hall-Press-2001* (Second Edi). Prentice Hall Press.

Song, Y. (2018). *Stock trend prediction: Based on machine learning methods*. UCLA.

Strader, T. J., Rozycki, J. J., ROOT, T. H., & Huang, Y.-H. (John). (2020). Machine Learning Stock Market Prediction Studies: Review and Research Directions. *Journal of International Technology and Information Management*, *28*(4), 63–83. https://doi.org/10.58729/1941-6679.1435

*Technical Analysis of Stocks and Trends Definition*. (n.d.). Retrieved March 19, 2023, from https://www.investopedia.com/terms/t/technical-analysis-of-stocks-and-trends.asp

Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock Closing Price Prediction using Machine Learning Techniques. *Procedia Computer Science*, *167*(2019), 599–606. https://doi.org/10.1016/j.procs.2020.03.326

Zhao, Y., Zhou, Y., & Lin, J. (2022). China Market Impact Cost Model Based on Machine Learning. *Procedia Computer Science*, *202*, 27–32. https://doi.org/10.1016/j.procs.2022.04.005

Zong, X. (2021). *Machine learning in stock indices trading and pairs trading*. University of Glasgow.

# APPENDIX

Some of the minute detail of the sections above can be relegated to the Appendix and referenced from the body of the report. Include all relevant documentation, computer coding, screen displays, etc.

# 