A Comparative Study of Machine Learning Algorithms for Trading Bot Performance

Siyabulela Monde Mathe

**ST10114635**

Postgraduate Diploma in Data Analytics

IN THE FACULTY OF INFORMATION TECHNOLOGY

AT VARSITY COLLEGE – NEWLANDS

2024 January

# Declaration

I hereby declare that the work presented in this thesis has not been submitted for any other degree or professional qualification, and that it is the result of my own independent work.

Siyabulela Monde Mathe

Full Name Goes Here (Candidate)

17/01/2024

Date

# Abstract

Abstract of about 300 words.

Machine learning algorithms are increasingly being used to develop trading bots. These algorithms can be used to identify patterns in financial data and to make trading decisions based on these patterns. However, there is a lack of research on the comparative performance of different machine learning algorithms for trading bot performance. This research will address this gap by conducting a comparative study of different machine learning algorithms for trading bot performance.

The performance of trading bots is affected by several factors, including the quality of the data, the choice of algorithm, the hyperparameters, and the trading strategy. By comparing the performance of different machine learning algorithms, this research will provide guidance on which algorithm is most likely to be successful in a particular trading environment.

This research is important because it will help to improve the performance of trading bots. By understanding which machine learning algorithms are most effective, researchers and practitioners can develop trading bots that are more likely to generate profits. This research will also benefit investors who are interested in using trading bots, as it will help them to choose the best algorithm for their needs.

# Acknowledgements

I would like to extend my sincere appreciation and gratitude to the following individuals and organizations who have played a pivotal role in the completion of this research:

Dr Rudolf Holzhausen: My lecturer and mentor who gave me unlimited amounts of guidance be it via class or templates even taking time out of his day to accommodate us with online classes. His faith in me has really kept me going.

# Table of contents

Contents

[Declaration i](#_Toc183110396)

[Abstract ii](#_Toc183110397)

[Acknowledgements ii](#_Toc183110398)

[Table of contents iii](#_Toc183110399)

[Chapter 1: Introduction 1](#_Toc183110400)

[1.1 Background and aims 1](#_Toc183110401)

[1.2 Thesis structure 2](#_Toc183110402)

[1.3 Contextualisation of the study 3](#_Toc183110403)

[1.4 The Research Questions 4](#_Toc183110404)

[1.4.1 Research Questions 4](#_Toc183110405)

[1.4.2 Aim 4](#_Toc183110406)

[1.4.3 Research Objectives 4](#_Toc183110407)

[1.5 Problem statement 5](#_Toc183110408)

[1.6 Assumptions and Rationale 6](#_Toc183110409)

[1.6.1 Assumptions: 6](#_Toc183110410)

[1.6.2 Rationales/Purpose Statement: 6](#_Toc183110411)

[1.7 Chapter Summary 7](#_Toc183110412)

[Chapter 2: Literature review 8](#_Toc183110413)

[2.1 Introduction 8](#_Toc183110414)

[2.2 Time Series Analysis 9](#_Toc183110415)

[2.3 What traditional approaches have researchers used in the past to understand financial assets? 10](#_Toc183110416)

[2.4 What are some of the drawbacks of existing works in the field of quantitative finance? 11](#_Toc183110417)

[2.5 Structured vs Unstructured Data 11](#_Toc183110418)

[2.6 Supervised Learning 12](#_Toc183110419)

[2.7 Unsupervised Learning 12](#_Toc183110420)

[2.8 AI vs ML vs DL 13](#_Toc183110421)

[2.8.1 Artificial Intelligence 13](#_Toc183110422)

[2.8.2 Machine Learning 14](#_Toc183110423)

[2.8.3 Deep Learning 15](#_Toc183110424)

[2.8.4 Commonly used AI powered Algorithms for Trading Bot Development 20](#_Toc183110425)

[2.8.5 Factors that Affect the Performance of Machine Learning Trading Bots: 21](#_Toc183110426)

[2.9 Algorithmic Trading 22](#_Toc183110427)

[2.10 Market Conditions 23](#_Toc183110428)

[2.11 Technical Analysis 24](#_Toc183110429)

[2.11.1 RSI – Relative Strength Index 26](#_Toc183110430)

[2.11.2 RSI and Time Series Analysis 27](#_Toc183110431)

[2.11.3 Support & Resistance/Demand(Buy low & Sell high) 28](#_Toc183110432)

[2.12 Fundamental Analysis 30](#_Toc183110433)

[2.13 Sentiment Analysis 31](#_Toc183110434)

[2.14 Chapter Summary 32](#_Toc183110435)

[Chapter 3: Methodology 33](#_Toc183110436)

[3.1 Introduction 33](#_Toc183110437)

[3.2 Research Method, Approach and Design 33](#_Toc183110438)

[3.2.1 Research Method 33](#_Toc183110439)

[3.2.2 Research Approach 34](#_Toc183110440)

[3.2.3 Research Design 35](#_Toc183110441)

[3.3 Data Collection 35](#_Toc183110442)

[3.3.1 Population 35](#_Toc183110443)

[3.3.2 Sampling 35](#_Toc183110444)

[3.3.3 Data Collection 36](#_Toc183110445)

[3.4 Data Analysis Methods 37](#_Toc183110446)

[3.4.1 Algorithms Used: 38](#_Toc183110447)

[3.5 Ethical Considerations 43](#_Toc183110448)

[3.5.1 Ethical web scraping 43](#_Toc183110449)

[3.6 Limitations of the study 44](#_Toc183110450)

[3.7 Chapter summary 45](#_Toc183110451)

[Chapter 4: Data Analysis 46](#_Toc183110452)

[4.1 Introduction 46](#_Toc183110453)

[4.2 4.2 Exploratory Data Analysis 46](#_Toc183110454)

[4.3 Data Pre- processing and preparation 49](#_Toc183110455)

[4.4 Model Implementation and Evaluation 50](#_Toc183110456)

[4.5 Training set evaluation 59](#_Toc183110457)

[4.5.1 Sell (1) 59](#_Toc183110458)

[4.5.2 Buy(2) 59](#_Toc183110459)

[4.6 Test set 59](#_Toc183110460)

[Chapter 5: Future Research and Conclusion 62](#_Toc183110461)

[5.1 Section 62](#_Toc183110462)

[References 62](#_Toc183110463)

[Chapter 6: References 62](#_Toc183110464)

[Figure 1 Flow chart Illustraying the structure of the conducted thesis. 2](#_Toc181629733)

[Figure 2 Time Series Analysis of NASDAQ Composite 8](#_Toc181629734)

[Figure 3 AI onion illustrating encompassing relation to ML AND DL 13](#_Toc181629735)

[Figure 4 Diagram illustrating evolotution and concept of neural networks 16](#_Toc181629736)

[Figure 5 Overview of Reinforcement Learning 19](#_Toc181629737)

[Figure 6 Diagram illustrating algo trading bots' logic 22](#_Toc181629738)

[Figure 7 Common chart patterns traders take into consideration when conducting technical analysis 24](#_Toc181629739)

[Figure 8 Diagram illustrating simulation of RSI signal 26](#_Toc181629740)

[Figure 9 Line chart illustrating key S/R levels 28](#_Toc181629741)

[Figure 10 Illustration of how fundamental analysis can affect price sentiment and trend in relation to foreign exchange currencry cross pairs e.g : EUR/USD 29](#_Toc181629742)

[Figure 11 Screenshot Illustrating EDA for Stock Data set used for sentiment analysis 45](#_Toc181629743)

[Figure 12 Illustrating EDA for EU cross pair data 46](#_Toc181629744)

[Figure 13 Illustrating Category distribution amongst EU dataset 47](#_Toc181629745)

[Figure 14 Illustrating EDA for NASDAQ composite price data obtain from yahoo finance API 48](#_Toc181629746)

[Figure 15 Screenshot illustrating the data preparation for sentiment analysis (removing punctuation and having uniform data/text) 48](#_Toc181629747)

[Figure 16 Importing the min and max scaler used to normalize the data 49](#_Toc181629748)

[Figure 17 Screenshot of data after the data was preprocessed through rough noise removal and scaling the data, removing outliers and normalization. 49](#_Toc181629749)

[Figure 18 Illustrating the train/test data split to avoid over and underfitting and initialization of the Gradient Booster EU model classifier 52](#_Toc181629750)

[Figure 19 Illustrating the Independent variables being fed to the model is the technical rsi indicatory, the three trend categories(0-no trend,1-downward trend and 2- the uptrend. With the dependent(target variable being the price movement). The price data such as open, close wasn’t necessary as this is meaningless data to the model and would only present noise to the data. 52](#_Toc181629751)

[Figure 20 Illustrating the initialization of the RNN model being built and initialized with 4 layers including and LSTM layer implemented to learn long-term dependencies 53](#_Toc181629752)

[Figure 21 Compilation of the model 53](#_Toc181629753)

[Figure 22 Illustrating predicted closing price vs test closing price of Nasdaq composite 54](#_Toc181629754)

[Figure 23 of visualisation illustrating time series analysis simulation/backtest of implemented S/R levels candlestick strategy and trade executions, along with backtesting performance metric results including drawdown. 54](#_Toc181629755)

[Figure 24 illustrating the backtest strategy performance metrics against the predefined strategy 56](#_Toc181629756)

[Figure 25 Illustrating Train and Test Accuracy Results of XGBoost Classifier 56](#_Toc181629757)

[Figure 26 Illustrating XGBoost Model Train Test Classification report 57](#_Toc181629758)

[Figure 27 Illustrating Test Classification Report of XGBoost classifier 57](#_Toc181629759)

[Figure 28 Illustration of the sentiment analysis classifiers classification matrix and confusion matrix used to evaluate the performance and accuracy of the model 58](#_Toc181629760)

[Figure 29 Table comparing the model performance metrics for both models including the XGBOOST EU classifier and the Multionomial Naive Bayes Sentiment Analysis classifier 59](#_Toc181629761)

# Introduction

## Background and aims

Price prediction, also known as price forecasting, is the process of estimating the future price of a commodity, product, or service based on historical data, current market conditions, and various other factors(Li et al,2021;Sezer et al, 2020). It involves analysing various factors that influence prices, such as supply and demand, economic indicators, competitor pricing, and seasonal trends(Li et al,2021;Sezer et al, 2020). Price prediction is used by businesses to make informed decisions about pricing strategies, inventory management, and resource allocation. It is also used by investors to make informed decisions about buying and selling financial assets.(Li et al,2021;Sezer et al, 2020)

There are two main approaches to price prediction. Technical analysis ultimately is the approach that focuses on analysing historical price charts and technical indicators to identify patterns and trends that can be used to predict future prices. Fundamental analysis focuses on analysing the underlying fundamentals of an asset, such as its financial performance, industry trends, and economic conditions, to predict its future price.

Both technical analysis and fundamental analysis can be used to make price predictions. However, they are best used together, as each approach can provide insights that the other may miss.

There are many factors that are commonly used in price prediction. Historical prices are one of the most important factors in price prediction. By analyzing historical price patterns, traders can identify trends and make informed decisions about future prices. Current market conditions, such as supply and demand, economic indicators, and competitor pricing, can also have a significant impact on future prices. Many commodities and products have seasonal trends in their prices. For example, the price of ice cream is typically higher in the summer months. News and events can also have a significant impact on future prices. For example, a major announcement about a new product or a change in government policy can cause prices to fluctuate wildly.

Price prediction is not an exact science. There is no one-size-fits-all approach that can guarantee accurate predictions(Li et al,2021;Sezer et al, 2020). However, by using a combination of technical analysis, fundamental analysis, and other factors, traders can increase their chances of making profitable trades.(Li et al,2021;Sezer et al, 2020)

.

## Thesis structure

A group of text on a table

Description automatically generated with medium confidence

Figure 1 Flow chart Illustraying the structure of the conducted thesis.

Chapter 1 focuses on the introduction and contextualisation of the study where I’ll be speaking about the problem, asking questions, showing the aim of this paper and any assumptions along the way.

Chapter 2 Focuses on the Literature Review where I am going to discuss the what traditional approaches have researchers used in the past to understand financial asset, what are some of the drawbacks of existing works in the field of quantitative finance, common ML algorithms used for trading bots, algorithmic trading and factors that ultimately affect/influence the performance of AI powered trading bots

.Chapter 3 focuses on the methodology, the data collection and gathering process.

Chapter 4 focuses on the analysis, here I’m presenting figures and their significance and as well as showing the performance of various machine learning algorithms.

Chapter 5 will be on the findings and the discussion of said findings.

Chapter 6 is the conclusion where I’ll conclude the study and give my recommendations for future research.

## Contextualisation of the study

Human traders are limited by their cognitive abilities and emotions. They can only process a finite amount of information and are susceptible to making mistakes due to fatigue, stress, or greed. Machine learning trading bots, on the other hand, are not limited by these constraints. They can process vast amounts of data and make decisions based on objective criteria. This makes them more likely to make accurate and profitable trading decisions. (Gemini,2024)

Machine learning trading bots can process data much faster than human traders. This allows them to identify trading opportunities and make trades more quickly. Supervised trading bots are not emotional. This means that they are not susceptible to making rash decisions based on fear or greed. Essentially reducing human errors. ML trading bots can be programmed to follow a specific trading strategy. This ensures that they are always making trades that are in line with their objectives. AI powered trading bots can be backtested and optimised. This means that their performance can be evaluated and improved over time. Machine learning algorithms can understand and comprehend quantitative financial data in ways that humans can’t.

## The Research Questions

### Research Questions

* What traditional approaches have researchers used in the past to understand financial assets?
* What are some of the drawbacks of existing works in the field of quantitative finance?
* Which machine learning algorithm has the best performance on a particular trading dataset with a particular quality?
* Are there any factors that affect the performance of machine learning trading bots, such as the trading strategy?
* What are the future challenges and opportunities for machine learning in trading bot development?

### Aim

The aim of this research is to critically assess, compare and evaluate various machine learning algorithms for trading bot performance

### Research Objectives

The objectives are:

* To identify the traditional approaches researchers have used in the past to understand financial assets.
* To identify and comprehend some of the drawbacks of existing works in the field of quantitative finance?
* To compare the performance of different machine learning algorithms for trading bot performance.
* To identify the factors that affect the performance of machine learning trading bots.
* To develop best practices for developing and deploying machine learning trading bots.
* To explore the future challenges and opportunities for machine learning in trading bot development.
* To raise awareness of the potential benefits and risks of using machine learning for trading.

## Problem statement

Human traders are limited by their cognitive abilities and emotions. They can only process a finite amount of information and are susceptible to making mistakes due to fatigue, stress, or greed. Machine learning trading bots, on the other hand, are not limited by these constraints. They can process vast amounts of data and make decisions based on objective criteria. This makes them more likely to make accurate and profitable trading decisions. Machine learning trading bots can process data much faster than human traders. This allows them to identify trading opportunities and make trades more quickly. Supervised trading bots are not emotional. This means that they are not susceptible to making rash decisions based on fear or greed. Essentially reducing human errors. ML trading bots can be programmed to follow a specific trading strategy. This ensures that they are always making trades that are in line with their objectives. Machine learning trading bots can be back tested and optimised. This means that their performance can be evaluated and improved over time. Machine learning algorithms can understand and comprehend quantitative financial data in ways that humans can’t.

## Assumptions and Rationale

### Assumptions:

* Data-Driven Insights: AI and ML can uncover hidden patterns and relationships in vast financial datasets, leading to superior investment strategies compared to traditional methods.
* Adaptive and Flexible: These algorithms can continuously learn and adapt to changing market dynamics, potentially outperforming static models based on historical data.
* Automation and Efficiency: AI and ML can automate cumbersome tasks like data analysis and portfolio optimization, freeing up quants for higher-level decision-making.
* Scalability and Generalizability: Algorithms trained on large datasets can be applied across diverse asset classes and markets, potentially generating consistent returns.

### Rationales/Purpose Statement:

The study will contribute to the existing knowledge of data science, artificial intelligence and machine learning in quantitative financial analysis and provide a deeper understanding of the impact of machine learning algorithms have on trading bot performance, and ultimately how the application of ML algorithms can be utilised for accurate algorithmic trading bot development. The research is expected to have a number of positive outcomes. Firstly, it is expected to identify the best machine learning algorithms for different types of trading bots and financial markets. This will help traders and developers to choose the right machine learning algorithm for their specific needs. Secondly, the research is expected to develop new machine learning algorithms that are specifically designed for trading bots. This will improve the performance of trading bots and make them more reliable and robust. Additionally, the research is expected to improve the understanding of how machine learning algorithms work in trading bots. This will help researchers and developers to develop better machine learning algorithms for trading bots in the future. Finally, the research is expected to develop new methods for evaluating the performance of trading bots. This will help traders and developers to make better decisions about which trading bots to use. The study will benefit various researchers in the field of algorithmic trading, FinTech stakeholders, investment Management Institutions, day traders/investors and banks. It will also benefit academic disciplines such as data science, quant science, and computer science. (Liu et al,2020; An et al,2022

## Chapter Summary

This chapter introduces the research topic of using machine learning algorithms for trading bot performance. It outlines the research questions and objectives, which aim to compare different algorithms and identify factors affecting their performance. The chapter also discusses the limitations of traditional approaches and the potential benefits of machine learning in trading.

# Literature review

## Introduction

AI powered algorithms are increasingly being used to develop trading bots. These algorithms can be used to identify patterns in large amounts of financial data and thus assist in making informed trading decisions based on these patterns. However, there is a lack of research on the comparative performance of different machine learning algorithms for trading bot performance (Liu et al,2020; An et al,2022)

. This literature review will discuss the different machine learning algorithms that have been used for trading bot development, the factors that contribute to the performance of machine learning trading bots, and the best practices for developing and deploying machine learning trading bots. (An et al, 2022; Rundo et al,2019)

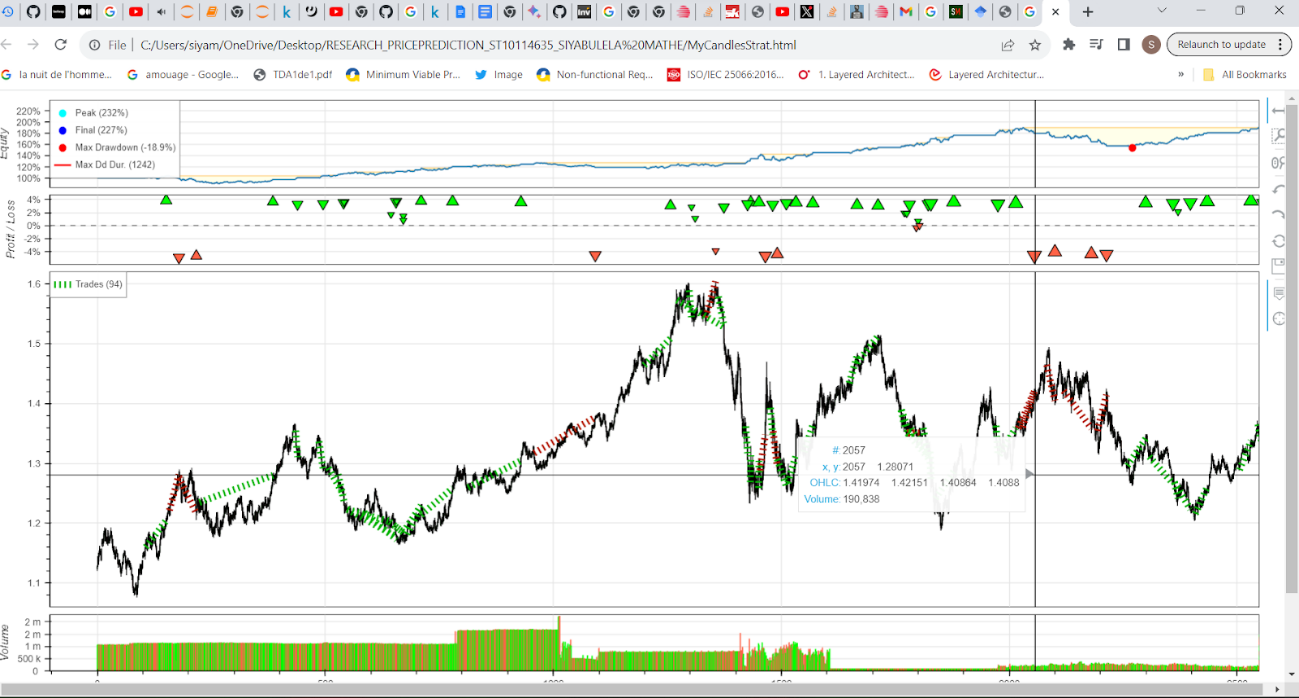


Figure 2 Time Series Analysis of NASDAQ Composite

## Time Series Analysis

In the scope of mathematics, a **time series** is a collection of indexed data points (or listed or graphed) in the order of time. Most commonly, a time series is a sequence retrieved at successive incremental time points(Jahan et al,2018). Thus, it is a sequence of discrete-time data. Examples of time series are heights of ocean tides, counts of sunspots, and the daily closing value of the Dow Jones Industrial Average. (Jahan et al,2018)

**A time series is a sequence of data points indexed in time order. It's often visualized as a line chart, showing how values change over time. Time series analysis is a valuable tool in various fields, including statistics, signal processing, finance, and engineering. It's used to identify trends, patterns, and anomalies in data, and to make predictions about future values.** (Jahan et al,2018)

**Time series *analysis*** comprises methods for analysing time series data with the primary aim of extracting meaningful statistics and other characteristics of the data Time series forecasting involves using a model to predict future values based on past data. This data is often treated as a random process(Jahan et al,2018). While regression analysis can be used to examine relationships between different time series, it's not typically considered time series analysis. True time series analysis focuses on relationships between data points within a single series over time. Time series data is unique because it's ordered by time. This differs from cross-sectional data, where the order of observations doesn't matter (like comparing people's wages based on education level). It also differs from spatial data, where observations are linked to specific locations (like analyzing house prices based on location and features).

When modeling time series data, we often assume that nearby data points are more closely related than distant ones(Jahan et al,2018). Additionally, time series models typically use a one-way flow of time, meaning that current values are influenced by past values, not future ones. This is known as time irreversibility. (Jahan et al,2018)

## What traditional approaches have researchers used in the past to understand financial assets?

Researchers have proposed several systems based on traditional approaches, such as the autoregressive integrated moving average (ARIMA) and the exponential smoothing model, in order to devise an accurate data representation. Linear models have also been suggested to perform time-series forecasting taking advantage of their effectiveness and quite simple implementation. Then, non-linear models (such as ARIMA) have drawn attention due to their ability to transform non-linear data into stationary ones. In statistics and economics, ARIMA and SARIMA models are used to analyze and predict time series data. These models are extensions of ARMA models, which assume that the data is stable over time.

**ARIMA** is used for time series data that has a trend. To handle this trend, the data is "differenced," which means calculating the difference between consecutive data points. This removes the trend and stabilizes the data.

**SARIMA** is used for time series data that has both a trend and seasonal patterns. To handle the seasonal pattern, the data is "seasonally differenced," which involves calculating the difference between data points at the same point in each season.

(Pricope,2021; Shavandi e al,2022)

## What are some of the drawbacks of existing works in the field of quantitative finance?

Existing works in the field of quantitative finance face some drawbacks due to poor performance when managing a large amount of data with deep-rooted complexity, high dimensionality, and casual dynamicity. Furthermore, traditional approaches such as the autoregressive integrated moving average (ARIMA) and the exponential smoothing model ***are not suitable for understanding hidden relationships (dependencies) between data.***

(An et al, 2022; Rundo et al,2019)

## Structured vs Unstructured Data

Structured data is organized and easily digestible for machines, making it readily accessible for analysis and insights. This type of data adheres to a predefined format, limiting its flexibility but simplifying its processing. Examples include numerical data like dates and phone numbers, as well as textual data like names and product descriptions. (Gemini,2024)

Unstructured data, on the other hand, lacks a predefined format, making it more challenging to analyse. While this flexibility allows for diverse data types like images, audio, and text, it requires more sophisticated techniques, often involving deep learning, to extract meaningful information. (Gemini,2024)

## Supervised Learning

In supervised learning, algorithms are trained on a dataset where each data point is paired with a corresponding label or output value(Alloghani,2020).This labeled data serves as a guide for the algorithm to learn a mapping function from input to output. (Alloghani,2020)

As the algorithm processes the data, it adjusts its internal parameters to minimize the difference between its predicted output and the actual label. This iterative process allows the algorithm to gradually improve its accuracy over time. Supervised learning is commonly used for tasks such as classification and regression.

**Price forecasting is primarily a supervised learning task.**

In supervised learning, we have a labeled dataset where each data point is associated with a corresponding output value. When applied to price forecasting, this means we have historical data with known prices for specific time periods(Alloghani,2020). We train a model on this data to learn the relationship between input features (e.g., historical prices, economic indicators, market sentiment) and the target variable (future price).

Once the model is trained, it can make predictions for future time periods based on new input data. This makes it a powerful tool for businesses and investors to make informed decisions.

## Unsupervised Learning

Unlike supervised learning, unsupervised learning involves training algorithms on unlabeled data. In this case, the algorithm does not have access to predefined output values. Instead, it must discover underlying patterns and structures within the data itself. By analyzing the data, the algorithm can identify similarities, differences, and anomalies, without any external guidance(Alloghani,2020). Common techniques in unsupervised learning include clustering and dimensionality reduction. Clustering involves grouping similar data points

together, while dimensionality reduction aims to reduce the number of features in a dataset while preserving important information. (Alloghani,2020)

## AI vs ML vs DL

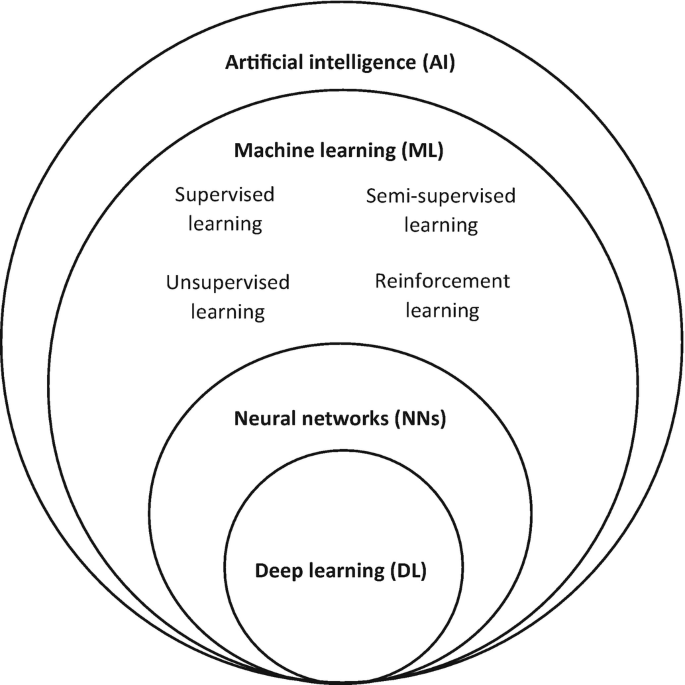


Figure 3 AI onion illustrating encompassing and cross secional relation to ML AND DL

### Artificial Intelligence

Artificial intelligence (AI) refers to the development of intelligent machines that can perceive their surroundings, learn from experience, and make decisions to achieve specific goals. It's a field of computer science focused on creating software and systems that mimic human intelligence(Chan et al, 2022). These intelligent machines, often simply referred to as AIs, are designed to interact with their environment and respond in ways that maximize their chances of success(Campesato,2023). Artificial intelligence is employed when a machine replicates human-like intelligence to complete a task. A good example is the Roomba vacuum cleaner, which utilizes AI to assess the room's size, identify obstacles, and determine the most efficient cleaning path, much like a human would. This demonstrates how AI can be used to automate and optimize tasks in our daily lives.

### Machine Learning

Machine learning (ML) is a subset of artificial intelligence that empowers computers to learn from data without explicit programming(Gemini,2024). By employing algorithms inspired by neural networks, statistics, and other disciplines, ML models can uncover hidden patterns and insights within data(Chan et al, 2022). This enables machines to independently learn and improve their performance on tasks like image recognition, natural language processing, and predictive analytics(Ongsulee,2017). For instance, ML can be used to train software to recognize specific objects in images or to predict future trends in stock prices Chan et al, 2022).

Regularization is a technique used in machine learning to prevent overfitting, a condition where a model becomes too complex and performs poorly on new, unseen data(Campesato,2023).. It achieves this by constraining or shrinking the model's coefficients, effectively reducing its complexity(Chan et al, 2022). By mitigating the impact of noise and irrelevant data points, regularization helps models generalize better and make more accurate predictions. (Ongsulee,2017)

Machine learning is ideal for tasks that involve teaching a model to recognize patterns or make predictions based on structured data. For instance, Spotify leverages machine learning to analyze your music preferences and the preferences of similar users, tailoring personalized playlists to your tastes. Chan et al, 2022)

By integrating machine learning with data analytics, businesses can gain valuable insights into market trends and consumer behaviour(Campesato,2023). This enables accurate forecasting of factors like demand, inventory needs, and transportation costs, leading to cost savings and improved operational efficiency(Chan et al, 2022). Machine learning, particularly well-suited for structured data, excels at identifying patterns and trends within large datasets, often uncovering insights that humans may miss. Unlike deep learning, which requires human intervention, machine learning models can autonomously learn and refine their predictions, reducing the need for manual effort. This automation not only saves time but also ensures reliable and consistent results. (Campesato,2023).

### Deep Learning

Deep learning, an advanced form of machine learning, leverages neural networks to analyze complex patterns within vast datasets(Ongsulee,2017). It's inspired by the human brain, aiming to replicate its ability to learn and recognize intricate patterns(Campesato,2023). Deep learning finds applications in challenging tasks like facial recognition, defect detection, and image processing, where traditional machine learning methods may fall short. (Chan et al, 2022)

Deep learning shines where machine learning struggles: handling unstructured data. From handwritten notes to images and voices, deep learning can unlock the hidden knowledge within this vast data source. (Chan et al, 2022)Deep learning's true strength lies in its scalability. It excels at processing massive datasets simultaneously and performing complex analyses at lightning speed. This translates to significant improvements in productivity, modularity, and portability for companies(Campesato,2023). Imagine running powerful deep neural networks on Google's Cloud AI platform, leveraging their infrastructure to scale batch predictions and optimize efficiency by automatically adjusting resources based on traffic demands. Additionally, deep learning models excel at utilizing parallel and distributed algorithms(Ongsulee,2017). This means the time it takes for a model to learn and refine itself is dramatically reduced. While training can be done locally on a single machine, when dealing with massive datasets, this becomes impractical(Chan et al, 2022). Parallel and distributed algorithms distribute the data (or the model) across multiple machines, making training more efficient and significantly speeding up the entire process, ultimately saving businesses valuable time and money. (Chan et al, 2022)

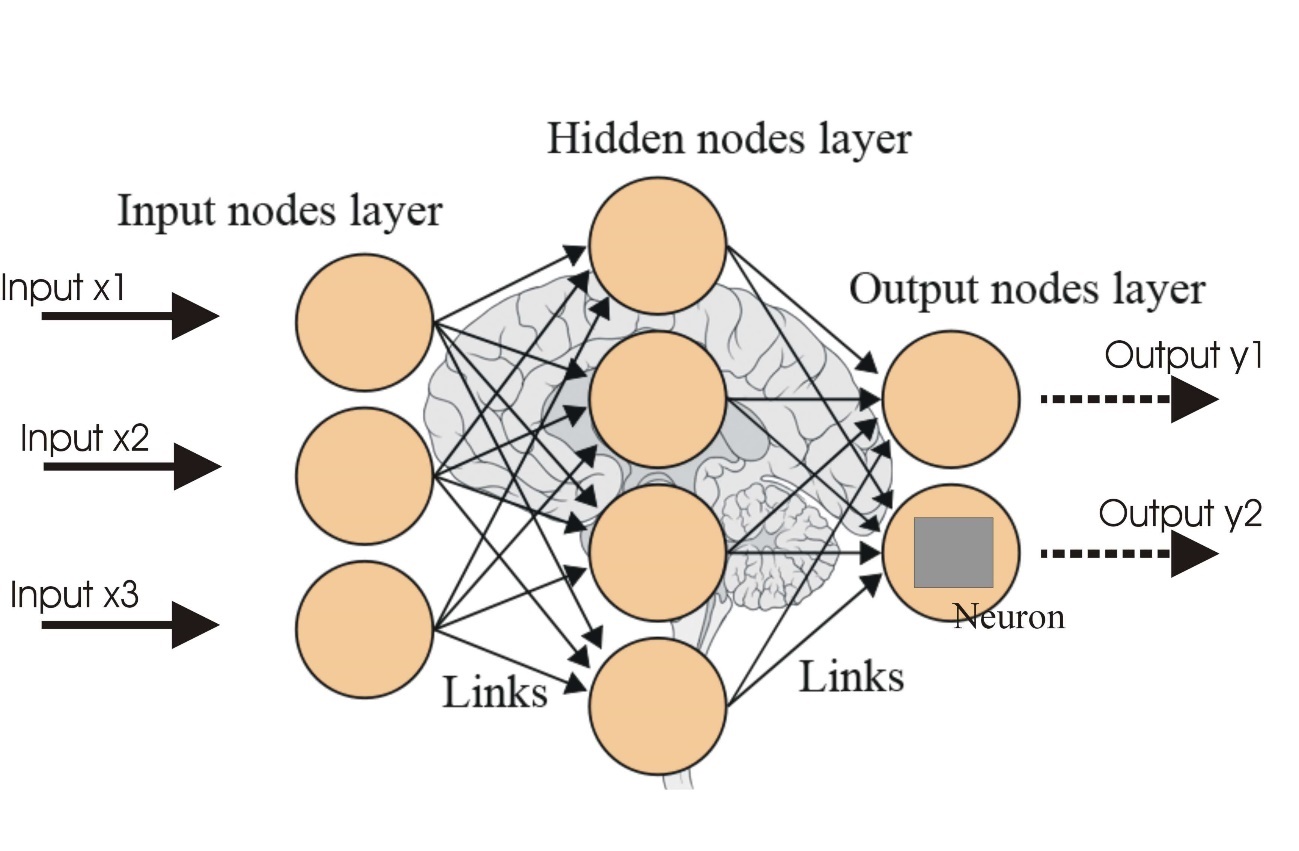


Figure 4 Diagram illustrating evolotution and concept of neural networks

#### Neural Networks

Neural networks, inspired by the human brain, are a type of machine learning model. They consist of interconnected nodes or artificial neurons, which process information through layers(Campesato,2023). These layers, often organized into input, hidden, and output layers, allow the network to learn complex patterns from data. As data flows through the network, it's adjusted by weights, which determine the strength of connections between neurons(Ongsulee,2017). Through this process, neural networks can learn to recognize patterns, make predictions, and solve complex problems, such as image recognition, natural language processing, and predictive modeling. They are particularly powerful when dealing with large and complex datasets. (Campesato,2023).

Neural networks are trained using a technique called empirical risk minimization. This involves adjusting the network's parameters to minimize the error between its predicted output and the actual correct output(Campesato,2023). Gradient-based methods, like backpropagation, are commonly used to fine-tune these parameters(Ongsulee,2017). During training, the network learns from labelled data, iteratively improving its predictions and reducing the error. This process enables the network to generalize its knowledge to new, unseen data. (Campesato,2023).

|  |  |  |
| --- | --- | --- |
| AI | ML | DL |
| AI mimics human intelligence to conduct tasks and make decisions. | Machine Learning is a subset of AI that utilizes algorithms to grasp patterns from data. | Deep Learning is a subset of ML that implements artificial neural networks for complex tasks. |
| AI may or may not require large datasets; it can use predefined rules. | ML requires labeled and unlabeled data for training and making predictions. | DL needs extensive labelled and unlabeled data and performs very well with big datasets. |
| Artificial Intelligence is considered rule-based, which ultimately means that it’s dependent human programming and intervention. | ML automates pattern recognintion and learning from data which ultimately means that it’s less dependent on manual intervention. | DL automates feature extraction, reducing the need for manual engineering. |
| AI is capable of handling numerous tasks, from simple to complex, across different domains. | ML excels in data-driven tasks such as classification, regression, etc. | DL specializes at tasks that are deemed very complex like image recognition, natural language processing, and more |
| AI algorithms can be simple or complex, depending on the application. | ML utilizes various supervised and unsupervised algorithms like decision trees, SVM, and random forests. | DL employs deep neural networks, which can have many hidden layers for complex learning. |
| AI may require less training time and resources for rule-based systems. | ML training time varies depending on the complexity of the task and size of the dataset. | DL training demands extensive computational resources and time dedicated for training deep neural networks. |
| AI systems may offer interpretable results based on human rules. | ML models can be interpretable or less interpretable based on the algorithm. | DL models are often considered less interpretable due to complex network architectures. |
| AI is mainly employed within the context of virtual assistants, recommendation systems, and more. | ML is largely used in image recognition, spam filtering, and other data tasks. | DL is utilized in autonomous vehicles, speech recognition, and advanced AI applications. |

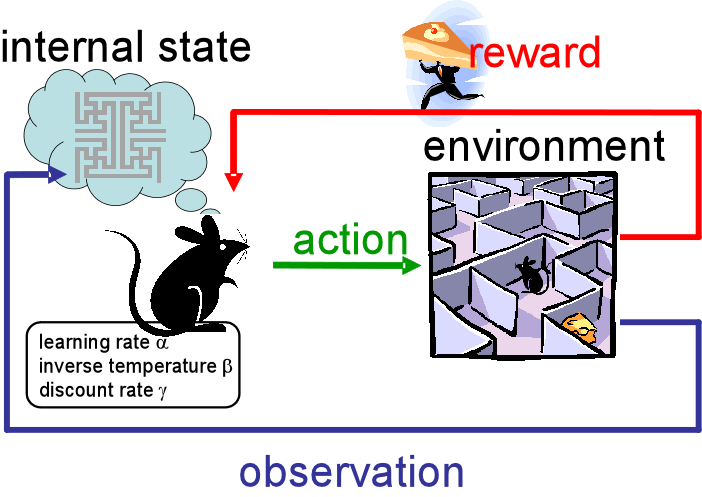


Figure 5 Overview of Reinforcement Learning

#### Reinforcement Learning

Reinforcement Learning Reinforcement learning (RL) is a type of machine learning that allows an agent to learn how to behave in an environment by trial and error. The agent is rewarded for taking actions that lead to desired outcomes and punished for taking actions that lead to undesired outcomes Pricope,2021; Shavandi e al,2022). Over time, the agent learns to take actions that maximise its rewards. RL can be utilised for the research topic in several ways. Firstly, RL can be used to develop trading bots that can learn to trade stocks or currencies profitably Pricope,2021; Shavandi e al,2022). The bot would be rewarded for taking actions that lead to profits and punished for taking actions that lead to losses. Over time, the bot would learn to trade in a way that maximises its profits. (Pricope,2021; Shavandi e al,2022)

### Commonly used AI powered Algorithms for Trading Bot Development

There are a variety of machine learning algorithms that can be used for trading bot development. Some of the most popular algorithms include. Support vector machines are a type of supervised learning algorithm that can be used to classify or regress data. SVMs have been shown to be effective for trading bot development, as they can be used to identify patterns in financial data and to make trading decisions based on these patterns(Gemini,2024). Decision trees are another type of supervised learning algorithm that can be used for trading bot development. Decision trees work by dividing the data into smaller and smaller subsets until the data can be classified or regressed. Decision trees have been shown to be effective for trading bot development, as they can be used to identify simple patterns in financial data. Neural networks are a type of machine learning algorithm that is inspired by the human brain. Neural networks can be used to learn complex patterns in data, and they have been shown to be effective for trading bot development. XGBoost is an open-source machine learning library that implements the gradient boosting algorithm. Gradient boosting is a machine learning technique that combines multiple weak learners to create a strong learner. XGBoost is a popular choice for trading bot performance because it is efficient as a fast and efficient algorithm that can be used to train large datasets. Can be considered highly accurate as it has been shown to be very accurate in a variety of trading applications. Highly is highly configurable, which allows users to fine-tune the algorithm to their specific needs.

(Liu et al,2020; An et al,2022)

### Factors that Affect the Performance of Machine Learning Trading Bots:

The quality of the data used to train a machine learning trading bot is critical to its performance. If the data is noisy or inaccurate, the machine learning bot will not be able to learn accurate patterns(An et al, 2022; Rundo et al,2019). The choice of machine learning algorithm is also important for the performance of trading bots. Some algorithms are better suited for certain types of data than others. The hyperparameters of a machine learning algorithm are the settings that control the behaviour of the algorithm(An et al, 2022; Rundo et al,2019). The hyperparameters can have a significant impact on the performance of the algorithm. The trading strategy that is used by the machine learning bot also affects its performance. The trading strategy should be designed to take advantage of the patterns that the machine learning bot has learned.(An et al, 2022; Rundo et al,2019)

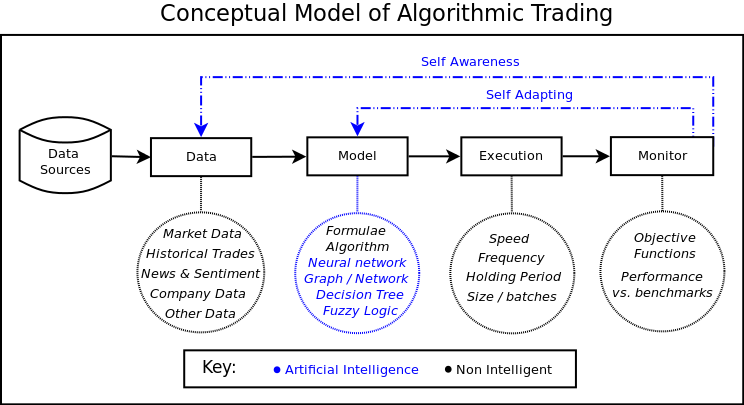


Figure 6 Diagram illustrating algo trading bots' logic

## Algorithmic Trading

Algorithmic trading leverages automated instructions to execute trades based on factors like time, price, and volume, taking advantage of computers' speed and computational power(Salkar et al, 2021). This method has gained significant popularity among both retail and institutional traders in the 21st century, with estimates according to Goldman Sachs suggesting that over 90% of Forex trading is now automated. (Barca,2017)

Algorithmic trading is employed by various financial institutions, including investment banks, pension funds, mutual funds, and hedge funds, to efficiently execute large orders or react swiftly to market changes(Barca,2017). However, it's also accessible to individual traders through retail tools.

The term algorithmic trading is often synonymous with automated trading systems, which encompass a variety of trading strategies(Barca,2017). These strategies can be based on mathematical models and rely on specialized software. Examples include systematic trading, market making, arbitrage, and high-frequency trading (HFT). HFT, in particular, involves rapid-fire trading with high turnover rates, utilizing advanced algorithms to execute trades based on real-time market data. (Barca,2017)

The rise of algorithmic trading and HFT has significantly transformed market microstructure and dynamics. It has altered the way liquidity is provided and introduced new complexities and uncertainties to the market. (Barca,2017)

Algorithmic trading is a type of trading that uses computer programs to execute trades. These programs are designed to follow specific rules or algorithms, and they can trade much faster than human traders (Pricope,2021; Shavandi e al,2022). Algorithmic trading is becoming increasingly popular, as it can help traders to improve their efficiency and profitability.

## Market Conditions

The performance of algorithmic trading strategies can be affected by market conditions. For example, algorithmic trading strategies that rely on technical analysis may perform better in trending markets, while strategies that rely on fundamental analysis may perform better in volatile and liquid markets. (An et al, 2022; Rundo et al,2019)

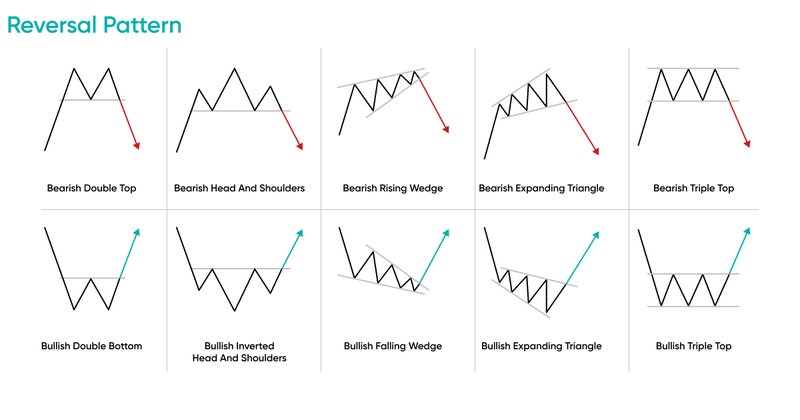


Figure 7 Common chart patterns traders take into consideration when conducting technical analysis

***Figure 7 illustrating the common chart patterns are very important as they provide traders with a key basis for technical analysis, as they are commonly hint that price is about to reverse off key S/R levels.***

## Technical Analysis

Technical analysis is a method of analysing financial market data to identify patterns that can be used to predict future price movements. Technical analysts use a variety of tools and indicators, such as moving averages, trendlines, and relative strength indexes, to identify these patterns. Technical analysts delve into the intricacies of price charts, searching for patterns like head and shoulders or double tops and bottoms. They meticulously study technical indicators, moving averages, and formations such as support and resistance lines, channels, flags, pennants, and cup and handle patterns. (Jahan et al,2018)

Additionally, they employ a wide range of market indicators, often mathematical transformations of price, volume, and other factors. These indicators help assess trends, their potential direction, and their likelihood of continuation. Analysts also explore the relationships between price/volume indices and market indicators, such as moving averages, relative strength index, and MACD. Other areas of interest include the correlation between options (implied volatility) and put/call ratios with price, as well as sentiment indicators like Put/Call ratios, bull/bear ratios, short interest, and implied volatility. (Salkar et al, 2021)

Technical analysis encompasses various techniques, each with its unique approach. Practitioners may focus on specific methods like candlestick analysis, harmonics, Dow theory, or Elliott wave theory, or combine elements from multiple techniques(Jahan et al,2018). While some analysts rely on subjective judgment to interpret patterns, others prefer a strictly mechanical or systematic approach.

Technical analysts believe that a market's price fully reflects all relevant information, including economic, fundamental, and news events(Jahan et al,2018). Rather than focusing on these external factors, they analyze historical price patterns of securities or commodities. The underlying assumption is that investor behavior is cyclical and tends to repeat itself, creating identifiable price trends and conditions. By studying these patterns, technical analysts aim to predict future price movements. (Jahan et al,2018)

Technical analysis is a popular tool for algorithmic trading, as it can be used to identify trading opportunities and manage risk. However, it is important to note that technical analysis is not a perfect science, and it cannot guarantee profits. (Pricope,2021; Shavandi e al,2022)

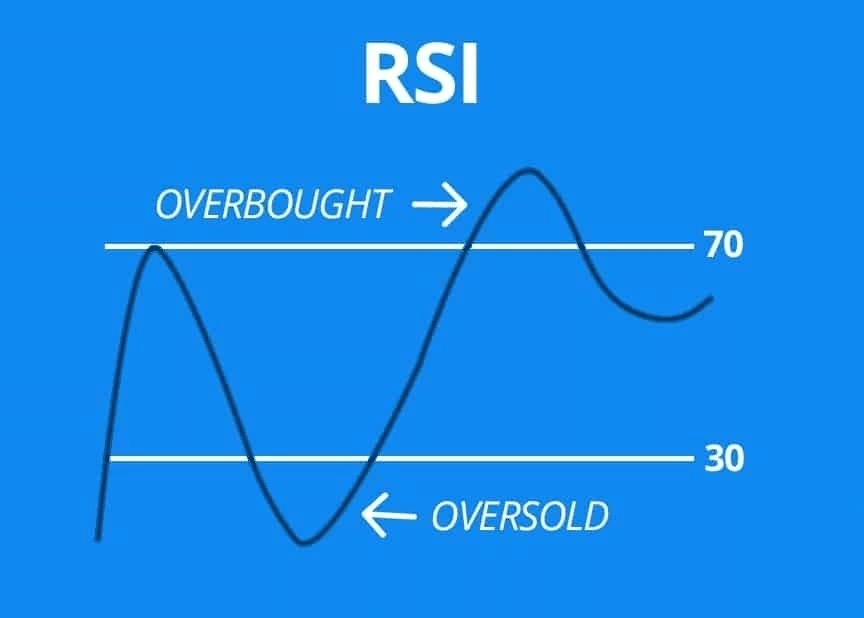


Figure 8 Diagram illustrating simulation of RSI signal

### RSI – Relative Strength Index

The relative strength index (RSI) is a technical indicator utilized in the analysis of financial data and markets. Its sole purpose is to chart the current and historical strength or weakness of a stock or market price which is derived from the closing prices of a recent trading period. The indicator shouldn’t be mistaken with relative strength. (Jahan et al,2018)

The RSI is defined as a momentum oscillator, which is responsible for measuring the velocity and magnitude of price movements. Momentum can be defined as the rate of the rise or fall in price(Yildririm et al,2019). The relative strength RS is given as the ratio of higher closes to lower closes. Concretely, one computes two averages of absolute values of closing price changes, i.e. two sums involving the sizes of candles in a candle chart. The RSI computes momentum as the ratio of higher closes to overall closes: stocks which have had more, or stronger positive changes have a higher RSI than stocks which have had more or stronger negative changes. (Jahan et al,2018)

The RSI is most typically used on a 14-day timeframe, measured on a scale from 0 to 100, with high and low levels marked at 70 and 30, respectively(Salkar et al, 2021). Short or longer timeframes are used for alternately shorter or longer outlooks. High and low levels—80 and 20, or 90 and 10—occur less frequently but indicate stronger momentum. (Yildririm et al,2019)

**How it Works:**

**Calculation:**

It calculates the average gain and average loss over a specific period (e.g., 14 days).

These averages are then used to compute the Relative Strength (RS).

The RSI is derived from the RS using a specific formula. (Yildririm et al,2019)

**Interpretation:**

**Overbought:** When the RSI rises above a certain level (typically 70), it suggests that the asset is overbought and may be due for a correction or pullback.

**Oversold:** When the RSI falls below a certain level (typically 30), it suggests that the asset is oversold and may be due for a rebound. (Salkar et al, 2021)

**Divergence:** When the price and RSI diverge, it can signal a potential trend reversal.

### RSI and Time Series Analysis

Time series analysis is the analysis of data points collected over time(Yildririm et al,2019). It involves identifying patterns, trends, and seasonality within the data. RSI, as a momentum oscillator, is closely related to time series analysis because:

**Momentum:** It measures the momentum of price changes, which is a key aspect of time series analysis.

**Trend Identification:** It helps identify trends, such as uptrends and downtrends, which are fundamental patterns in time series data.

**Cycle Detection:** It can be used to identify cyclical patterns in price movements, which is a common characteristic of many time series. (Yildririm et al,2019)

**Predictive Modeling:** While RSI itself is not a predictive model, it can be used as a feature in more complex time series models to improve forecasting accuracy.

**In essence, RSI provides a valuable tool for understanding the dynamic nature of price movements over time**(Yildririm et al,2019)**. By incorporating RSI into time series analysis, traders and analysts can make more informed decisions about buying, selling, or holding assets.**

### ****Support & Resistance/Demand(Buy low & Sell high)****

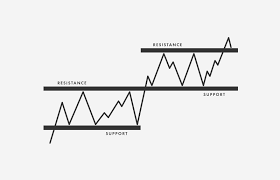


Figure 9 Line chart illustrating key S/R levels

Within financial markets technical analysis, support and resistance are specific predetermined zones of the price of an asset at which it is considered that the price will tend to stop and reverse. These levels are denoted by multiple touches of price without a breakthrough of the level. (Yildririm et al,2019)

#### Support versus resistance

A support level is a zone where the price tends to identify support as it drops due to an increase in demand for the asset. Essentially this results in the price being more likely to "reject" off this zone as opposed to than breaking through it(Yildririm et al,2019). However, once the price of the asset has broken through this level, by an amount exceeding some noise, it’s most is likely possible that it will to proceed in dropping until it meets another support level

A resistance level is the exact opposite of a support zone. It can essentially be considering a region in price, where the price is inclined to finding resistance as it rises due to an increase in selling interest(Yildririm et al,2019). Again, this ultimately means that the price is more inclined to "reject" off this level rather than break through it. However, once the price has broken this level, by an amount exceeding some noise, it is likely to continue rising until meeting another resistance level.

Traders utilized support and resistance zones in various chart patterns.[[3]](https://en.wikipedia.org/wiki/Support_and_resistance#cite_note-3)

#### Reactive versus proactive support and resistance

Support and resistance levels are crucial concepts in technical analysis, used to predict potential price reversals and identify potential entry and exit points. Proactive support and resistance levels are predicted based on current price action and technical analysis tools like Measured Moves, Fibonacci Retracements, Pivot Points, and Trendlines(Yildririm et al,2019). Reactive support and resistance levels, on the other hand, form directly as a result of price action and volume, such as Price Swing Highs/Lows and Volume Profile. Additionally, psychological levels, often round numbers, can act as significant support and resistance levels due to their psychological impact on traders(Jahan et al,2018). The more a level is tested and respected by the market, the stronger it becomes. A break above resistance or below support can often lead to a reversal of roles, with the broken level becoming the new support or resistance.

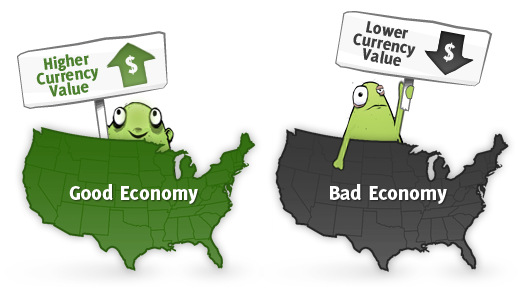


Figure 10 Illustration of how fundamental analysis can affect price sentiment and trend in relation to foreign exchange currencry cross pairs e.g : EUR/USD

## Fundamental Analysis

Fundamental Analysis Fundamental analysis is a method of analysing a shares's value by examining underlying factors such as a company's financial statements, economic data, and industry trends.

It is a long-term investment strategy that aims to identify undervalued or overvalued securities. Essentially, fundamental analysis entails news that can impact the trend of a market. (Pricope,2021; Shavandi e al,2022) In the context of the stock exchange, fundamental analysis can be used to identify companies that are undervalued based on their financial performance. For example, a company with strong earnings growth and a low price-to-earnings ratio may be considered undervalued. Fundamental analysis can also be used to identify industries that are expected to grow in the future. For example, an industry that is benefiting from technological innovation may be considered undervalued. In the context of the foreign exchange market, fundamental analysis can be used to identify currencies that are undervalued or overvalued based on economic factors such as interest rates, inflation, and economic growth. For example, a currency with a high interest rate may be considered undervalued if the country's economy is growing at a healthy pace(Liu et al,2020; An et al,2022). Fundamental analysis can also be used to identify countries.currencies that are expected to experience economic growth in the future. For example, a country with a young and growing population may be considered undervalued. (Pricope,2021; Shavandi e al,2022). Fundamental analysis can influence this research topic in a number of ways. Firstly, fundamental analysis can be used to identify stocks or currencies that are undervalued or overvalued. This information can be used to develop trading strategies that are based on the expectation that the prices of these securities will move in a certain direction. Secondly, fundamental analysis can be used to identify market trends. This information can be used to develop trading strategies that are based on the expectation that the market will continue to move in a certain direction. (Pricope,2021; Shavandi e al,2022) Overall, fundamental analysis can be a valuable tool for traders who are looking to make informed investment decisions. It can be used to identify undervalued or overvalued securities, identify market trends, and develop trading strategies that are based on these factors. (An et al, 2022; Rundo et al,2019)

## Sentiment Analysis

Sentiment analysis, also known as opinion mining or emotion AI, is a technique that uses natural language processing, text analysis, and computational linguistics to identify and analyze human sentiment within text data(Liu et al,2020). This can range from simple positive or negative sentiment to more complex emotions like anger, joy, or sadness. It's widely applied to customer reviews, social media posts, and healthcare records to gain insights into customer satisfaction, brand perception, and patient sentiment(Liu et al,2020. Recent advancements in deep language models, like RoBERTa, have expanded the capabilities of sentiment analysis, allowing for the analysis of more complex texts, such as news articles, where sentiment may be less explicit. (An et al,202

Sentiment analysis is the process of extracting subjective information from text, such as opinions, emotions, and beliefs. It can be used to analyse a variety of text sources, such as news articles, social media posts that are related to news, and customer reviews. In the context of the stock exchange and foreign exchange market, sentiment analysis can be used to identify the overall mood of investors or traders. This information can be used to develop trading strategies that are based on the expectation that the prices of stocks or currencies will move in a certain direction. (Liu et al,2020; An et al,2022) For example, if sentiment analysis shows that investors are bullish (optimistic) about the stock market, then this could be interpreted as a signal that stock prices are likely to rise. Conversely, if sentiment analysis shows that investors are bearish (pessimistic) about the stock market, then this could be interpreted as a signal that stock prices are likely to fall. News can also influence sentiment analysis. For example, if a news article reports that a company has released positive earnings results, then this could lead to a positive sentiment among investors(An et al, 2022; Rundo et al,2019). Conversely, if a news article reports that a country's central bank has raised interest rates, then this could lead to a negative sentiment among investors. Sentiment analysis can influence the discussed topic in a number of ways. First, sentiment analysis can be used to identify news articles that are likely to have a positive or negative impact on the sentiment of investors. This information can be used to develop trading strategies that are based on the expectation that the prices of stocks or currencies will move in a certain direction(An et al, 2022; Rundo et al,2019). Secondly, sentiment analysis can be used to identify patterns in the sentiment of investors over time. This information can be used to develop trading strategies that are based on the expectation that the prices of stocks or currencies will continue to move in a certain direction. (An et al, 2022; Rundo et al,2019) In conclusion, sentiment analysis can be a valuable tool for traders who are looking to make informed investment decisions. It can be used to identify the overall mood of investors or traders, identify news articles that are likely to have a positive or negative impact on the sentiment of investors, and identify patterns in the sentiment of investors over time.

## Chapter Summary

Chapter 2 delves into the theoretical foundations of time series analysis, machine learning, and deep learning. It explores traditional approaches like ARIMA and their limitations. The chapter also discusses the potential of machine learning algorithms, particularly neural networks, in improving trading bot performance. It highlights the importance of technical analysis, fundamental analysis, and sentiment analysis in providing valuable insights for trading strategies.

# Methodology

## Introduction

## This chapter outlines the research methodology employed to investigate the effectiveness of machine learning algorithms in predicting stock price movements. A quantitative research approach was adopted, utilizing a comparative design to assess the performance of various algorithms. The study involved a multi-step process, including data collection, pre-processing, feature engineering, model training, and evaluation. The methodological framework was designed to ensure the rigor and validity of the research findings.

## Research Method, Approach and Design

### Research Method

Positivism. The discussed research is part of this research paradigm because it is concerned with the objective measurement of the performance of different machine learning algorithms. Positivism is a research paradigm that assumes that there is a single reality that can be objectively measured and studied. The researcher in this case is trying to measure the performance of different machine learning algorithms by comparing their returns on investment, evaluation matrixes and maximum drawdowns. This is an objective measure of the performance of the algorithms, and it is consistent with the positivist research paradigm. (Bezuidnhout et al,2014;Liu et al,2020)Positivists believe that the researcher should be objective and impartial in their research. This means that the researcher should not let their own personal biases or beliefs influence the results of their research. Positivists believe that the best way to understand the world is to measure it. This means that they use quantitative methods, such as surveys and experiments, to collect data. Positivists believe that the goal of research is to predict the future. This means that they are interested in finding relationships between variables that can be used to make predictions. The research that’s going to be conducted is consistent with all of these characteristics of the positivist research paradigm. I am trying to be objective and impartial in my research, through using quantitative methods to collect data, and ultimately trying to predict/measure the future performance of machine learning algorithms.(Bezuidnhout et al,2014;Liu et al,2020)

### Research Approach

Quantitative. The research discussed essentially is a statistical method that enables researchers to collect and analyse quantitative/numerical data. This type of data is well-suited for the research topic because it is objective and can be easily measured. Quantitative analysis can be used to:

* Collect data: Quantitative analysis can be used to collect data from a variety of sources, such as surveys, experiments, and historical records.
* Analyse data: Quantitative analysis can be used to analyse data using statistical methods, such as regression analysis, correlation analysis, and hypothesis testing.
* Interpret data: Quantitative analysis can be used to interpret data and draw conclusions about the relationships between variables.
* Make predictions: Quantitative analysis can be used to make predictions about future events based on the relationships between variables. (Bezuidnhout et al,2014;Liu et al,2020)

Here are some of the benefits of using quantitative analysis for the research topic:

* It is objective: Quantitative analysis is based on numerical data, which is objective and can be easily measured. This makes it a more reliable method of research than qualitative methods, which are based on subjective data, such as interviews and focus groups.
* It is generalizable: The results of quantitative analysis can be generalized to a larger population, making it a more powerful method of research than qualitative methods.
* It is efficient: Quantitative analysis can be used to collect and analyze large amounts of data quickly and easily. This makes it a more efficient method of research than qualitative methods. (Bezuidnhout et al,2014;Liu et al,2020)

### Research Design

Considering the algorithms were tested on historical data, the following research designs/strategies were implemented:

* A comparative study: which involved comparing the performance of different machine learning algorithms on a variety of secondary financial datasets.
* A simulation study: This would involve using a computer simulation to test the performance of different machine learning algorithms. (Bezuidnhout et al,2014;Liu et al,2020)

## Data Collection

### Population

The study will benefit various researchers in the field of algorithmic trading, FinTech stakeholders, investment Management Institutions, day traders/investors and banks. (Pricope,2021; Shavandi e al,2022).

### Sampling

A simple random sample or a stratified random sample was considered the best option. This is because the goal of the study was to compare the performance of different machine learning algorithms, and a random sample will ensure that the results of the study are representative of the population of machine learning algorithms.(Bezuidenhout et al,2014;Liu et al,2020)However, if the we were only interested in a particular type of machine learning algorithm, then a purposive sample may be a better option. For example, if the researcher is only interested in algorithms that use technical analysis, then they could select a purposive sample of algorithms that use technical analysis.(Bezuidenhout et al,2014;Liu et al,2020)

### Data Collection

The first step was to collect secondary data using yahoo finance’ yfinance python package(API) where NASDAQ Composite historical data and EUR/USD currency cross pair will be collected.(Densmore,2017;Bezuidnhout et al,2014) Web scraping is the process of extracting data from websites using automated means. APIs, or Application Programming Interfaces, are tools that allow software applications to communicate with each other. By using APIs, web scrapers can extract data from websites without having to manually visit each page.(Densmore,2017;Bezuidnhout et al,2014)

Historical data is data that has been collected over time. It can be used to track trends, identify patterns, and make predictions. In the context of finance, historical data can be used to track the performance of stocks, bonds, and other financial instruments. Web scraping historical data using APIs can be a useful way to collect large amounts of data quickly and easily. (Densmore,2017;Bezuidnhout et al,2014)

This data can then be used for a variety of purposes, such as:

● Analysing market trends: Historical data can be used to analyse market trends, such as price movements and trading volumes. This information can be used to identify potential trading opportunities.

● Identifying patterns: Historical data can be used to identify patterns in market prices. This information can be used to develop trading strategies that are based on these patterns.

● Historical data can be used to make predictions about future market prices. This information can be used to make informed investment decisions. However, it is important to note that web scraping historical data using APIs can be challenging. This is because websites are constantly changing, and the APIs that are used to access them can also change. As a result, it is important to regularly update the web scraper to ensure that it is still working properly.

Here are some of the benefits of using web scraping historical data using APIs:

● It is a quick and easy way to collect large amounts of data.

● It is a reliable way to collect data, as the data is directly from the source.

● It is a cost-effective way to collect data, as there are no fees associated with using APIs. (Densmore,2017;Bezuidnhout et al,2014)

\* \*\*Data collection:\*\* The first step is to collect data on the data using investing.coms’ investpy python package(API) where NASDAQ Composite historical data and GBP/USD currency cross pair will be collected. We’ll aim to collect at least 36 months of supervised data\* (Densmore,2017;Bezuidnhout et al,2014)

\*\*Data preprocessing:\*\* The data that you collect will need to be preprocessed before it can be used for machine learning. This involves cleaning the data and removing any errors or outliers. (Pricope,2021; Shavandi e al,2022) \*

\*\*Model selection:\*\* Once the data has been preprocessed, you need to select the machine learning algorithm that you will use for your study. There are a variety of machine learning algorithms that can be used for trading bot development, and the choice of algorithm will depend on the specific research questions that you are asking. (An et al, 2022; Rundo et al,2019) \*

\*\*Model training:\*\* Once you have selected the machine learning algorithm, you need to train the model on the historical data. This involves feeding the data to the algorithm and allowing it to learn the patterns in the data. (Pricope,2021; Shavandi e al,2022) \*\*Model evaluation:\*\* Once the model has been trained, you need to evaluate its performance. This can be done by backtesting the model on historical data or by deploying the model in a live trading environment.(An et al, 2022; Rundo et al,2019) \*\*Model Optimization:

\*\* Models will be optimised and fine tuned using hyper parameter tuning. (Bezuidnhout et al,2014;Liu et al,2020)

## Data Analysis Methods

The study primarily involvde time series analysis which is a subset of **predictive** data analysis. (Gemini,2024)

Here's why:

**Predicting Future Outcomes(time series forecasting):** The primary aim is to forecast future stock prices or market trends using historical data. This is a classic predictive analytics task.

**Model Building:** Machine learning algorithms are used to build models that can identify patterns and relationships in the data, allowing for predictions of future price movements.

**Decision Making:** The predictions generated by these models can inform trading decisions, such as when to buy or sell a particular asset.

While the study might involve some elements of **prescriptive** analysis, such as determining optimal trading strategies based on predicted outcomes, the primary focus remains on prediction

### Algorithms Used:

#### XGBoost classification

XGBoost, or eXtreme Gradient Boosting, is a powerful and versatile open source machine learning library that excels in various tasks like classification, regression, and ranking(Gemini,2024). It operates on the principle of gradient boosting, which involves sequentially adding weak models (often decision trees) to create a stronger ensemble model. XGBoost optimizes this process through techniques like regularization, parallel processing, and efficient handling of missing data(Gemini,2024). It's renowned for its scalability and accuracy, making it a popular choice for both small-scale and large-scale machine learning projects. XGBoost can be deployed on single machines or distributed across clusters, enabling it to handle massive datasets and complex models. (Gemini,2024)

#### Multinomial Naïve Bayes

Naive Bayes is a straightforward technique for building classifiers that assign class labels to data points. It operates on the assumption that features are independent of each other, given the class label. This means that each feature contributes independently to the probability of a particular class. For instance, a fruit might be classified as an apple based on its color, shape, and size, with each feature contributing independently to the overall probability. (Gemini,2024)

While Naive Bayes is often associated with Bayesian probability, it can also be used with maximum likelihood estimation, a more frequentist approach. This allows for practical applications without fully embracing Bayesian methods. (Gemini,2024)

The Multinomial Naive Bayes model is a statistical method for classifying data, particularly text documents. It assumes that each feature (word) in a document is independent of other features, given the document's class. This model represents documents as histograms, where each word's frequency is a feature. The probability of a document belonging to a particular class is calculated based on the frequencies of words in the document and their associated probabilities in the training data. (Gemini,2024)

To improve the model's robustness, a technique called Laplace smoothing is often employed. This involves adding a small value (pseudocount) to all frequency counts, preventing zero probabilities and ensuring that all features contribute to the classification decision. By using log-space calculations, the model mitigates potential numerical issues associated with multiplying many small probabilities.

Multinomial Naïve Bayes for Sentiment Analysis

A Multinomial Naive Bayes classifier, combined with a Bag-of-Words approach, was a popular technique for sentiment analysis. Here's a breakdown of how it worked:

**Data Collection:**

**News Articles:** News articles, press releases, and social media posts related to a specific stock or the broader market were gathered.

**Financial Reports:** Quarterly and annual financial reports, analyst reports, and earnings call transcripts were collected.

**Text Preprocessing:**

**Cleaning:** Stop words, punctuation, and irrelevant information were removed.

**Tokenization:** Text was broken down into individual words or tokens.

**Stemming/Lemmatization:** Words were reduced to their root form to group similar words together.

**Bag-of-Words Model:**

Each document (news article, report) was represented as a bag of words, where each word was a feature and its frequency was the feature value.

For example, the sentence "This stock is a great buy" would have been represented as:

{great: 1, buy: 1, stock: 1, is: 1, this: 1}

**Multinomial Naive Bayes Classifier:**

**Training:** The classifier was trained on a labeled dataset where each document was assigned a sentiment label (positive, negative, or neutral).

**Prediction:** Given a new document, the classifier calculated the probability of the document belonging to each sentiment class based on the frequency of words in the document and their associated probabilities in the training data.

**Key Considerations:**

**Feature Engineering:**

**N-gram Features:** N-grams (sequences of n words) were considered to capture context and sentiment nuances.

**Sentiment Lexicons:** Sentiment lexicons (e.g., SentiWordNet) were incorporated to assign sentiment scores to words.

**Model Evaluation:**

Metrics like accuracy, precision, recall, and F1-score were used to evaluate the model's performance.

Cross-validation was used to assess the model's generalization ability.

**Limitations:**

The Bag-of-Words model ignored word order, which can be important for understanding sentiment.

It struggled to capture complex sentiment expressions and sarcasm.

#### Recurrent Neural Network

**Recurrent Neural Networks (RNNs)** are a type of artificial neural network specifically designed to process sequential data. Unlike traditional neural networks, which treat each input independently, RNNs have a "memory" component that allows them to consider the context of previous inputs.  (Jahan et al,2018)

This "memory" is implemented as a hidden state, which is updated at each time step based on the current input and the previous hidden state(Jahan et al,2018). This enables RNNs to capture dependencies between data points in a sequence, making them well-suited for tasks like:

**Natural Language Processing (NLP):**

Text generation

Machine translation

Sentiment analysis

Text summarization

**Speech Recognition:**

Converting spoken language into text

**Time Series Analysis:**

Stock price prediction

Weather forecasting

**Music Generation**

Recurrent Neural Networks (RNNs) have their roots in both statistical mechanics and neuroscience. Early inspiration came from the Ising model and Hebbian learning, which led to the development of Hopfield networks. Neuroscience, particularly the study of the brain's recurrent neural circuits, also influenced the development of RNNs. In the 1980s, researchers like Jordan and Elman applied RNNs to cognitive psychology. However, training deep RNNs was challenging due to the vanishing gradient problem. (Gemini,2024)To address this, Schmidhuber introduced the concept of self-supervised pre-training and the neural history compressor. Later, he and Hochreiter developed Long Short-Term Memory (LSTM) networks, which effectively solved the vanishing gradient problem and became a cornerstone of modern RNN architectures. In parallel, inspired by statistical mechanics, researchers developed various architectures for unsupervised learning, including Boltzmann machines, restricted Boltzmann machines, and Helmholtz machines(Gemini,2024). These models aimed to learn deep generative models from unlabeled data. (Jahan et al,2018)

## Ethical Considerations

### Ethical web scraping

Ethics of Web Scraping of public data:

One should make usage of a public API when available and avoid scraping all together if the data you’re looking for is available through the API. Pass your data through a user agent string to identify who you are. Scrape data at a reasonable rate and throttle/control the number of requests per second. The website owner must not think it is a DDoS attack. Make sure your enterprise saves only the data it needs. Don’t scrape private data – Look at the site’s robots.txt and analytics needs to avoid scraping data from sensitive areas.

Data ownership: The data that is scraped may be owned by the website or organization that hosts it. Scraping this data without permission could be considered a violation of their intellectual property rights.

Privacy: Scraping data could also violate the privacy of the individuals or organizations that are the subject of the data. This is especially true if the data includes personal information, such as names, addresses, or phone numbers.

Botnets: Web scraping can be used to create botnets, which are networks of computers that are controlled by a single entity. Botnets can be used for malicious purposes, such as sending spam or launching cyberattacks. (Jahan et al,2018)

Market manipulation: Web scraping can also be used to manipulate markets. For example, a bot could be used to buy or sell large quantities of shares in a company, in order to artificially inflate or deflate the price of the stock.

(Densmore,2017;Bezuidnhout et al,2014)

It is important to be aware of these ethical issues when conducting web scraping for research purposes. By taking steps to mitigate these risks, researchers can help to ensure that their research is conducted in an ethical manner.

Here are some tips for conducting ethical web scraping:

Get permission: If possible, get permission from the website or organization that hosts the data before scraping it.

Be transparent: Be transparent about your research and why you are scraping the data.

Use a bot that is respectful of the website's resources: Use a bot that does not overload the website or disrupt its operation.

Only scrape public data: Do not scrape data that is considered private or confidential.

Use the data for good: Use the data for research purposes that are beneficial to society.

## Limitations of the study

The researcher's own biases could influence the way they design the study or interpret the results. The participants' own biases could also influence the results of the study. The Hawthorne effect is a phenomenon where participants in a study change their behaviour simply because they are being studied. This could impact the results of the study. The research topic is broad and could be interpreted in many different ways. It is important to clearly define the scope of the research and to focus on a specific set of issues. Quantitative analysis can be complex and time-consuming to learn and use. The results of quantitative analysis are limited by the quality and quantity of the data that is available. Quantitative analysis can be insensitive to the context of the data, which can lead to misleading results. Quantitative analysis requires large amounts of data to be effective. This can be a challenge in some cases, such as when the data is not readily available or when it is difficult to collect. The quality of the data used in quantitative analysis can also affect the results. If the data is inaccurate or incomplete, the results of the analysis may be misleading. Quantitative analysis often relies on making assumptions about the data. If these assumptions are not met, the results of the analysis may be inaccurate. Quantitative analysis can be complex and time-consuming. This can make it difficult to implement and interpret the results. The results of quantitative analysis can be difficult to interpret. This can be a challenge when trying to communicate the results to others. (Bezuidnhout et al,2014;Liu et al,2020)

## Chapter summary

This chapter outlines the research methodology, including the data collection process, data analysis techniques, and ethical considerations. The focus is on using a quantitative approach to analyse historical financial data and evaluate the performance of different machine learning models. The chapter also discusses the challenges and limitations of the study, such as data quality and model overfitting.

# Data Analysis

## Introduction

This chapter highlights the approach deployed to build a classification model with the aim of predicting price movements. The methodology involves the usage of supervised machine learning for dataset classification. The process starts with the collection of the dataset, followed by data pre- processing, feature selection implementation, model training and testing, and the subsequent execution of the classification algorithms.

## 4.2 Exploratory Data Analysis

A screenshot of a computer

Description automatically generated

Figure 11 Screenshot Illustrating EDA for Stock Data set used for sentiment analysis

A screenshot of a computer

Description automatically generated

Figure 12 Illustrating EDA for EU cross pair data

Checking the distribution of category classes in our data and it’s safe to say no trend is the least and classes one and two are the dominating classes in our data which is a good thing because those classes indicate a clear trend (down or up) which is needed to make trading predictions and decisions.

A screenshot of a computer

Description automatically generated

Figure 13 Illustrating Category distribution amongst EU dataset

While exploring the data, the intention was ultimately to get a visual understanding of what I was working with in the dataset, and which were the prevalent classes in our dataset. Interestingly enough, the dataset has more buy categories than sell predictions. This will be of use later when testing and evaluating our model.

A screenshot of a computer

Description automatically generated

Figure 14 Illustrating EDA for NASDAQ composite price data obtain from yahoo finance API

## Data Pre- processing and preparation

A screenshot of a computer

Description automatically generated

Figure 15 Screenshot illustrating the data preparation for sentiment analysis (removing punctuation and having uniform data/text)

A screenshot of a computer

Description automatically generated

Figure 16 Importing the min and max scaler used to normalize the data

A screenshot of a computer

Description automatically generated

Figure 17 Screenshot of data after the data was preprocessed through rough noise removal and scaling the data, removing outliers and normalization.

A white screen with black text

Description automatically generated

Figure 18 illustrating the backtest strategy performance metrics against the predefined strategy

## Model Implementation and Evaluation

The Recurrent Neural Network NASDAQ composite price prediction model was evaluated using the performance metrics of the simulation of the predefined backtested strategy in which the metrics measured potential profitability and risk.

These metrics included:

* **Start:** The starting date of the backtest period.
* **End:** The ending date of the backtest period.
* **Duration:** The total length of the backtesting period.
* **Exposure Time:** The percentage of the total time the strategy was actively invested.
* **Equity Final:** The final equity value at the end of the backtest period.
* **Equity Peak:** The highest equity value reached during the backtest period.
* **Return:** The total percentage return generated by the strategy.
* **Buy & Hold Return:** The return of a simple buy-and-hold strategy over the same period.
* **Volatility (Ann.):** Annualized volatility of the strategy's returns.
* **Sharpe Ratio:** Risk-adjusted return, measuring excess return per unit of risk.

**Risk Metrics:**

* **Max. Drawdown:** The maximum percentage decline in equity from peak to trough.
* **Avg. Drawdown:** The average percentage decline in equity during drawdown periods.
* **Max. Drawdown Duration:** The duration of the longest drawdown period.
* **Avg. Drawdown Duration:** The average duration of drawdown periods.

**Profitability Metrics:**

* **Win Rate:** The percentage of winning trades.
* **Best Trade:** The largest profit from a single trade.
* **Worst Trade:** The largest loss from a single trade.
* **Avg. Trade:** The average profit or loss per trade.
* **Max. Trade Duration:** The longest duration of a single trade.
* **Avg. Trade Duration:** The average duration of trades.
* **Profit Factor:** The ratio of gross profits to gross losses.
* **Sortino Ratio:** Similar to the Sharpe Ratio, but penalizes downside volatility more heavily.
* **Calmar Ratio:** Ratio of annualized return to maximum drawdown.

The strategy was backtested with a principal amount of $10, 000. The final equity of was indeed $22652.55 after an execution of 94 trades. The win rate was 75% with the best trade successfully generating 4 percent of the account size. The backtested strategey accumulated a return of 126.52%

The peak of the accounts equity was $23, 157.25 , however this was reduced due to the losing rate being 25% with the worst trade depleting 5% of the accounts equity

The XGBoost and Multinomial Naïve Bayes Sentiment Analysis Models were evaluated using classification reports and a confusion matrix.

Positivism is a philosophical approach that emphasizes empirical evidence and scientific methods. It suggests that knowledge is derived from observation and experimentation.

Quantitative Analysis is a research method that involves collecting and analysing numerical data. It uses statistical techniques to describe, explain, predict, and control phenomena.

Model Evaluation using classification reports and confusion matrices aligns with both positivism and quantitative analysis in the following ways:

* Empirical Evidence: Both classification reports and confusion matrices provide empirical evidence of a model's performance. They quantify the model's accuracy, precision, recall, and F1-score based on actual data.
* Objectivity: These metrics are objective and can be independently verified. They are not subjective interpretations but rather concrete measurements of the model's ability to correctly classify instances.
* Statistical Rigor: The underlying calculations and statistical tests used to derive these metrics adhere to rigorous statistical principles.
* Quantitative Nature: The metrics themselves are numerical, making them amenable to quantitative analysis and comparison.
* Falsifiability: The models can be tested against new data, and if they fail to perform as expected, their predictions can be refuted. This aligns with the principle of falsifiability, a cornerstone of scientific inquiry.

In essence, using classification reports and confusion matrices to evaluate models is a positivist and quantitative approach to understanding and improving the performance of machine learning algorithmic based trading bots. It provides a solid foundation for making data-driven decisions and drawing reliable conclusions.

A screenshot of a computer

Description automatically generated

Figure 19 Illustrating the train/test data split to avoid over and underfitting and initialization of the Gradient Booster EU model classifier

A screenshot of a computer

Description automatically generated

Figure 20 Illustrating the Independent variables being fed to the model is the technical rsi indicatory, the three trend categories(0-no trend,1-downward trend and 2- the uptrend. With the dependent(target variable being the price movement). The price data such as open, close wasn’t necessary as this is meaningless data to the model and would only present noise to the data.

A screenshot of a computer

Description automatically generated

Figure 21 Illustrating the initialization of the RNN model being built and initialized with 4 layers including and LSTM layer implemented to learn long-term dependencies

A screenshot of a computer

Description automatically generated

Figure 22 Compilation of the model

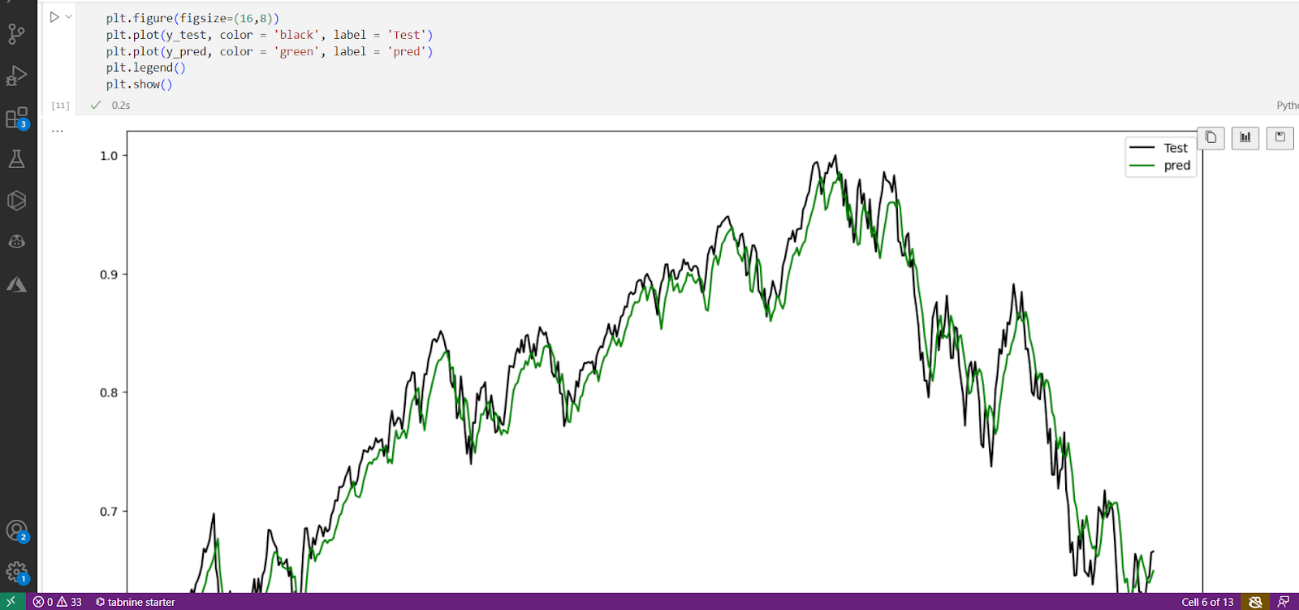
Figure illustrating neural network closing price predictor for NASDAQ composite

Figure 23 Illustrating predicted closing price vs test closing price of Nasdaq composite

A screen shot of a computer screen

Description automatically generated

Figure 24 of visualisation illustrating time series analysis simulation/backtest of implemented S/R levels candlestick strategy and trade executions, along with backtesting performance metric results including drawdown.

Backtesting is a technique used to evaluate the performance of a trading strategy or predictive model on historical data. It involves applying the model to past data and comparing its predicted outcomes to the actual results. (Vezeris et al,2020)

Why is Backtesting Important in Time Series Analysis?

In the context of time series analysis for trading, backtesting is crucial for several reasons:

* Reality Check: It provides a realistic assessment of how a model would have performed in the past, under real-world conditions.
* Risk Assessment: By testing the model on historical data, you can identify potential risks and limitations.
* Optimization: It allows you to fine-tune the model's parameters and hyperparameters to optimize its performance.
* Overfitting Prevention: Backtesting helps to identify overfitting, where a model performs well on training data but poorly on new, unseen data.
* Strategy Validation: It validates the effectiveness of a trading strategy by testing it against historical market data.

  (Vezeris et al,2020)

A screenshot of a computer

Description automatically generated

Figure 25 illustrating the backtest strategy performance metrics against the predefined strategy

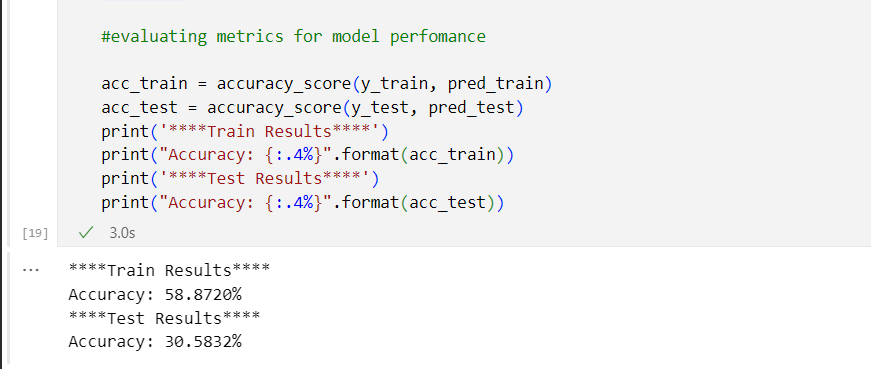


Figure 26 Illustrating Train and Test Accuracy Results of XGBoost Classifier

We are fitting the model on our training set then we’re asking the model to try predict the target on the training set and it’s providing 58 percent accuracy then for the test results we have 30 percent of accuracy. This is very bad as we can’t make clear price decisions of 30 percent accuracy.

Secondly, is that the 30 percent accuracy, we don’t know how it’s distributed and executed between the three category classes, it’s a global classification result between the 0,1,2 categories however the 0 class isn’t helpful because it’s a no trend classification hence why we don’t necessary consider it when making trading decision and why we’ll delve deep deeper with the classification reports.

A screenshot of a computer

Description automatically generated

Figure 27 Illustrating XGBoost Model Train Test Classification report

A screenshot of a computer

Description automatically generated

Figure 28 Illustrating Test Classification Report of XGBoost classifier

Figure Illustrating Test Classification Report of XGBoost classifier

## Training set evaluation

### Sell (1)

The downtrends are predicted within 59 percent precision, we are mostly interested in the test set but the, and a recall of 33 percent, meaning we are only detecting 53 percent of cases that present in the data, we are not very sensitive, but we are precise for the downtrend. Overall, the f1 score which englobes precision and recall is 54 percent. At some point we have to sacrifice the recall in trading, we can’t detect every single opportunity in the market, this is impossible, if you have something that is between 20 to 30 percent it’s okay. The most important thing is the precision. When you detect something it should be a true positive, thus correctly predicting the trend

### Buy(2)

The uptrend category we have 58 percent precision, we have 74 percent recall, meaning 74 of our predictions are correct, we have a very sensitive model to buying positions, this is huge as it is the prevalent class(category in our data and model)

## Test set

The buying model outperforms our sell model clearly due to the fact that it’s the most prevalent class, which is the pitfall to machine learning. Apart from the fact that the model 80 percent of the time has no clear direction/prediction of the market these figures are simply not good enough for trading and making investment decisions.

A screenshot of a computer

Description automatically generated

Figure 29 Illustration of the sentiment analysis classifiers classification matrix and confusion matrix used to evaluate the performance and accuracy of the model

The model is capable of accurately predicting market sentiment and direction by the headlines obtained from the dataset and the classification reports figures for example 93 percent for positive(buy as 0) and 80 percent for negative(sell as 1), the models is showing that it’ll be able to generalise to new data points and headlines with ease as the models accuracy was 85 percent.

|  |  |  |
| --- | --- | --- |
| Metric | XG Boost EU classifier | Multinomial NB Sentiment Analysis classifier |
| Accuracy | 31% | 85% |
| Precision(Buy) | 34% | 93% |
| Recall(Buy) | 61% | 76% |
| F – 1 Score(Buy) | 44% | 84% |
| Precision(Sell) | 24% | 80% |
| Recall(Sell) | 34% | 95% |
| F – 1 Score(Sell) | 28% | 87% |

Figure 30 Table comparing the model performance metrics for both models including the XGBOOST EU classifier and the Multionomial Naive Bayes Sentiment Analysis classifier

Both models performed quite differently in terms of how they were able to generalize new data points. The multinomial naive bayes sentiment analysis classifier was able to outperform the gradient booster classifier in both buy and sell classes, by producing remarkably higher precision values(93 and 80 percent respectively). This ultimately means that the sentiment analysis classifier can detect 93 percent of actual true positive buys and 80 percent of true positive sells. Thus being able to detect 80 percent of overall trades, suggesting it’s a very accurate model as opposed to the gradient booster classifier which had produced rather disappointing figures associated to generalizing new data points for buys and sells(34 and 24 percent respectively). This ultimately suggests that the gradient booster classifier can detect 34 percent of actual true positive buys and 24 percent of true positive sells, which is very underwhelming. These figures cannot be used in forecasting market decisions/bias.

# Future Research and Conclusion

## Section

The results indicated that classifiers underperformed well in identifying clear trading decisions. What we can conclude is that we for future development, we should build different models focused on primary goal, one for buying and another for selling, because the more categories, data, the more staff you ask of your models the more noise present in the data, thus making it harder to predict. We might have higher chances of success if we include smaller time frame simulations and modern technical liquidity-based indicators. The logical next step would be to test additional machine learning models to find an optimal one. By incorporating more intricate network architectures like a Generative Artificial Neural Network or even a Deep Reinforcement Learning agent.

Sentiment analysis was very accurate using the combination of naïve bayes and bag of words approach. It produced quiet accurate figures. However analysing market sentiment for the day will only get us the overall direction and bias of the day and not technical execution of the trade. Combining both may help for future research. A bot that follows an algorithm that obtains direction through market sentiment for the day then executes off a technical indicator parameter and only has one feature depending on the sentiminet/direction for the day.

# References

# References

Pricope, T.V., 2021. Deep reinforcement learning in quantitative algorithmic trading: A review. arXiv preprint arXiv:2106.00123.

Rundo, F., Trenta, F., di Stallo, A.L. and Battiato, S., 2019. Machine learning for quantitative finance applications: A survey. Applied Sciences, 9(24), p.5574.

An, B., Sun, S. and Wang, R., 2022. Deep reinforcement learning for quantitative trading: Challenges and opportunities. IEEE Intelligent Systems, 37(2), pp.23-26

Liu, X.Y., Yang, H., Chen, Q., Zhang, R., Yang, L., Xiao, B. and Wang, C.D., 2020. FinRL: A deep reinforcement learning library for automated stock trading in quantitative finance. arXiv preprint arXiv:2011.09607.

Shavandi, A. and Khedmati, M., 2022. A multi-agent deep reinforcement learning framework for algorithmic trading in financial markets. Expert Systems with Applications, 208, p.118124.

Densmore, J. 2017. Ethics in web scraping. Towards Data Science. <https://towardsdatascience.com/ethics-in-web-scraping-b96b18136f01>

Bezuidenhout, R., Davis, C. and Plooy-Cilliers, F. du (eds.) (2014) Research matters. Lansdowne, South Africa: Juta Legal and Academic.

Vuong, P.H., Dat, T.T., Mai, T.K. and Uyen, P.H., 2022. Stock-price forecasting based on XGBoost and LSTM. Computer Systems Science & Engineering, 40(1).

Sheng, L., 2021, August. Stock Market Movement Prediction: A Comparative Study Between Machine Learning and Deep Time Series Models. In International Conference on Computing and Data Science (pp. 15-27). Singapore: Springer Nature Singapore.

Sezer, O.B., Gudelek, M.U. and Ozbayoglu, A.M., 2020. Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. Applied soft computing, 90, p.106181.

Islam, S.F.N., Sholahuddin, A. and Abdullah, A.S., 2021. Extreme gradient boosting (XGBoost) method in making forecasting application and analysis of USD exchange rates against rupiah. In Journal of Physics: Conference Series (Vol. 1722, No. 1, p. 012016). IOP Publishing.

Li, Y., Xie, Y., Yu, C., Yu, F., Jiang, B. and Khushi, M., 2021. Feature importance recap and stacking models for forex price prediction. arXiv preprint arXiv:2107.14092.

Jahan, I. and Sajal, S., 2018. Stock price prediction using recurrent neural network (RNN) algorithm on time-series data. In *2018 Midwest instruction and computing symposium*. Duluth, Minnesota, USA: MSRP.

Lee, J. W. (2001). Stock price prediction using reinforcement learning. In Industrial Electronics, 2001. Proceedings. ISIE 2001. IEEE International Symposium on (Vol. 1, pp. 690-695). IEEE.

Barca Linares, V., 2017. Algorithmic Trading Technology and Strategy Research on Financial Markets.

Campesato, O., 2020. *Artificial intelligence, machine learning, and deep learning*. Mercury Learning and Information.

Ongsulee, P., 2017, November. Artificial intelligence, machine learning and deep learning. In *2017 15th international conference on ICT and knowledge engineering (ICT&KE)* (pp. 1-6). IEEE.

Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A. and Aljaaf, A.J., 2020. A systematic review on supervised and unsupervised machine learning algorithms for data science. *Supervised and unsupervised learning for data science*, pp.3-21.

Van, Q.T., 2023, December. Combining Machine Learning with Support Resistance Method in a Trading Strategy. In *2023 RIVF International Conference on Computing and Communication Technologies (RIVF)* (pp. 336-341). IEEE.

Chan, J.Y.L., Phoong, S.W., Cheng, W.K. and Chen, Y.L., 2022. Support resistance levels towards profitability in intelligent algorithmic trading models. *Mathematics*, *10*(20), p.3888.

Yıldırım, E.O., Uçar, M. and Özbayoğlu, A.M., 2019, November. Evolutionary optimized stock support-resistance line detection for algorithmic trading systems. In *2019 1st International Informatics and Software Engineering Conference (UBMYK)* (pp. 1-6). IEEE.

Salkar, T., Shinde, A., Tamhankar, N. and Bhagat, N., 2021, June. Algorithmic trading using technical indicators. In *2021 International Conference on Communication information and Computing Technology (ICCICT)* (pp. 1-6). IEEE.

Vezeris, D.T., Schinas, C.J., Kyrgos, T.S., Bizergianidou, V.A. and Karkanis, I.P., 2020. Optimization of backtesting techniques in automated high frequency trading systems using the d-Backtest PS method. *Computational Economics*, *56*, pp.975-1054.

* Gemini AI. (2024).

Introducing Gemini: our largest and most capable AI model. Google Blog. <https://blog.google/technology/ai/google-gemini-ai/>

(Vezeris et al,2020)

(Salkar et al, 2021)

(Yildririm et al,2019)

(Chan et al, 2022)

(Van, 2023)

(Campesato,2023)

(Ongsulee,2017)

(Alloghani,2020)

(Barca,2017)

(Jahan et al,2018)

(Bezuidnhout et al,2014;Liu et al,2020)

(Densmore,2017;Bezuidnhout et al,2014)

(An et al, 2022; Rundo et al,2019)

(Liu et al,2020; An et al,2022)

(Pricope,2021; Shavandi e al,2022)