AI–Assisted Value Investment in Malaysian Stock Market with Generative AI

BY

TEO TECK WAN

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF COMPUTER SCIENCE (HONOURS)

Faculty of Information and Communication Technology

(Kampar Campus)

JAN 2024

UNIVERSITI TUNKU ABDUL RAHMAN

REPORT STATUS DECLARATION FORM

Title: AI–Assisted Value Investment in Malaysian Stock Market with Generative AI

Academic Session: Jan 2024

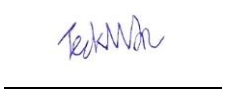
I TEO TECK WAN

(CAPITAL LETTER)

declare that I allow this Final Year Project Report to be kept in
Universiti Tunku Abdul Rahman Library subject to the regulations as follows:

1. The dissertation is a property of the Library.
2. The Library is allowed to make copies of this dissertation for academic purposes.

Verified by,



(Author's signature)

(Supervisor's signature)

Address:
No 16, Lorong Dato Abdul Hamid 16A,
Taman Sentosa, 41200 Klang,
Selangor.

Prof Ts Dr. Liew Soung Yue
Supervisor's name

Date: 23 April 2024

Date: 26/4/2024

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

UNIVERSITI TUNKU ABDUL RAHMAN

Date: 23 April 2024

SUBMISSION OF FINAL YEAR PROJECT /DISSERTATION/THESIS

It is hereby certified that Teo Teck Wan (ID No: 20ACB01750) has completed this final year project entitled "AI–Assisted Value Investment in Malaysian Stock Market with Generative AI" under the supervision of Prof Ts Dr. Liew Soung Yue (Supervisor) from the Department of Computer Science, Faculty of Information and Communication Technology.

I understand that University will upload softcopy of my final year in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,

(*TEO TECK WAN*)

*Delete whichever not applicable

DECLARATION OF ORIGINALITY

I declare that this report entitled "**METHODOLOGY, CONCEPT AND DESIGN OF A 2-MICRON CMOS DIGITAL BASED TEACHING CHIP USING FULL-CUSTOM DESIGN STYLE**" is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.

Signature : 

Name : TEO TECK WAN

Date : 23 April 2024

ACKNOWLEDGEMENTS

I would like to express my sincere thanks and appreciation to my supervisors, Prof. Ts. Dr Liew Soung Yue who has given me this bright opportunity to engage in a stock market and large language model related project. It is my first step to establish a career in large language model field. A million thanks to you.

Next, I would like to say thanks to my friends for their patience, unconditional support, and love, and for standing by my side during hard times. Finally, I must say thanks to my parents and my family for their love, support, and continuous encouragement throughout the course.

ABSTRACT

This project is an analysis-assisted tool for the stock market, offering human-like investment advice along with supportive information to validate the recommendations for users. The prediction tool in the stock market, which relies on quantitative data such as stock prices, volume, dividends, etc., is already well-established in the industry for forecasting stock prices. However, the approach to leveraging qualitative data for stock market analysis is still in its nascent stages. With the recent trends in AI, products like ChatGPT, Google Gemini, and Claude 3 are significantly influenced user interactions and decision-making processes. Given the increasing recognition of its capabilities across various industries, our team has chosen to harness ChatGPT, a leading AI product, to analyse qualitative data. We have customized ChatGPT to better suit the financial sector, enabling it to provide numerical scores along with logical, fact-based justifications for the short, medium, and long-term prospects of companies. This makes our tool exceptionally valuable to a wide range of users, from stock market professionals to everyday investors. The benefits of using our tool are substantial, including significant time and cost savings. For example, novice investors can save on the costs and time typically spent on stock market courses and expert consultations, while professionals can streamline their fundamental analysis of companies and market trends. Thus, this project introduces an alternative method to enhance investment decision-making.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

GPT	Generative Pre-trained Transform
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
DSS	Decision Support System
HMM	Hidden Markov Model
NN	Neural Networks
RNN	Recurrent Neural Network
SVM	Support Vector Machines
SVR	Support Vector Regression
NLP	Natural Language Processing
LLM	Large Language Model
EMH	Efficient Market Hypothesis
AMH	Adaptive Market Hypothesis
AI	Artificial Intelligence
GDP	Gross Domestic Product
CPI	Consumer Price Index
GNN	Graph Neural Networks
EKF	Extended Kalman Filter
VAR	Vector Autoregression
k-NN	k-Nearest Neighbors
RSI	Relative Strength Index
LSTM	Long-Short-Term Memory
CHAID	Chi-square Automatic Interaction Detectors
CART	Classification and Regression Trees
GARCH	Generalized Autoregressive Conditional Heteroscedasticity
E-GARCH	Exponential-GARCH
DAN2	Dynamic Artificial Neural Network
MLP	Multi-Layer Perceptron
RCNN	Recurrent Convolutional Neural Network
EPCNN	Evolving Partially Connected Neural Network
RBFNN	Radial Basis Function Neural Network
AFSA	Artificial Fish Swarm Algorithm

GA	Genetic Algorithms
PSO	Particle Swarm Optimization
ABC	Artificial Bee Colony
SRCS	Stepwise Regression-Correlation Selection
GRU	Gated Recurrent Units
ILDA	Interdependent Latent Dirichlet Allocation
EMA	Exponential Moving Average
ANFIS	Adaptive Network-based Fuzzy Inference System
CPDA	Cumulative Probability Distribution Approach
RST	Rough Set Theory
EMRLF	Evolutionary Morphological-Rank-Linear
MGA	Modified Genetic Algorithm
MRL	Morphological Rank Linear
LMS	Least Mean Square
POS	Part-of-Speech
GPOMS	Google Profile of Mood States
CRSP	Center for Research in Security Prices
MAE	Mean Absolute Error
MSE	Mean Square Error
RMSE	Root Mean Square Error
MAPE	Mean Absolute Percentage Error
CRISP-DM	Cross-Industry Standard Process for Data Mining
AWS	Amazon Web Services
RDS	Relational Database Service
API	Application Programming Interface
PDF	Portable Document Format

Chapter 1

Introduction

1.1 Problem Statement and Motivation

Recently, inflation has significantly impacted the global economy, gradually altering our everyday living conditions. A clear example is the rising cost of daily necessities, including food and beverages. Consequently, relying solely on an average income from a single job often falls short of covering daily expenses. Therefore, investing in the stock market presents an opportunity to supplement income. However, haphazard investments without thorough research can lead to financial losses. To enhance their chances of selecting stocks that will appreciate in value, individuals typically adopt two approaches: self-learning investment strategies or seeking the expertise of a fund manager. If opting for self-learning, individuals must invest time in stock market training courses and consider the costs associated with expert-led classes. On the other hand, hiring a fund manager to oversee investments typically involves a substantial fee. However, it might be worthwhile to explore finding a balance or even lower between the time and costs of self-learning and the fees associated with hiring a fund manager. This balanced approach might offer a more efficient and economical way to enhance investment skills.

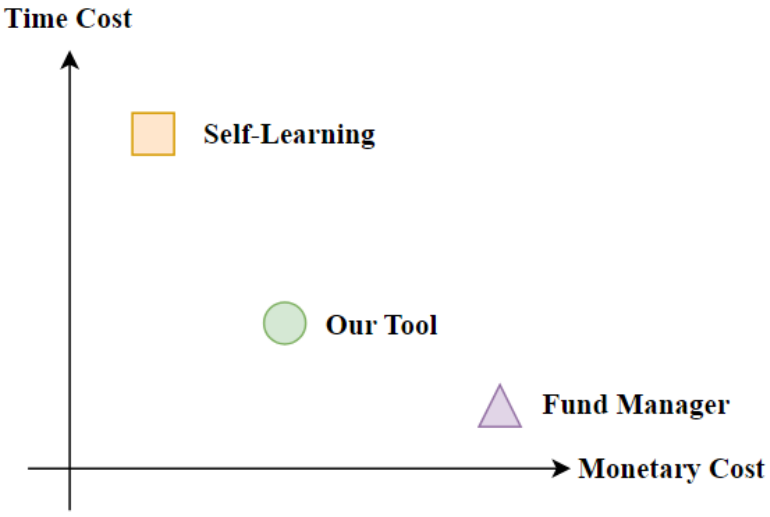


Figure 1.1. Visualization of the Costs associated with Related Approaches

Rather than studying the stocks and becoming experts or looking for fund manager, we are exploring simpler approaches for beginners in the stock market to select suitable

stocks for them. One of the simpler methods is a stock market prediction or analysis tool. These tools are many and available in the stock market, and they use various techniques like technical analysis and machine learning to forecast stock prices. They serve as mature aids for both experts and newcomers in making stock-buying decisions. The majority of current tools employ quantitative data for forecasting, which depends on historical information to discern past trends and patterns as a basis for future projections. However, quantitative data not always accurate to predict the future events despite the trends and patterns are identified in the past, because stock markets can be influenced by unpredictable variables like economic changes, political events, and so on. Hence, the role of qualitative data such as company news, government's policies or statement by companies' directors used for fundamental analysis, which predicting the future based on their planning and credibility.

In summary, the goal of this project is to develop a tool or system that enhances investment decision-making by leveraging qualitative information, such as company strategies and market insights, while ensuring our analyses are fact-based and well-founded. This tool aims to offer reliable stock market analysis suitable for both novice and experienced investors. The ultimate objective is to assist individuals in effectively analysing the stock market and delivering appropriate recommendations to our users.

1.2 Objectives

The project is aptly titled "AI-Assisted Value Investment in the Malaysian Stock Market with Generative AI" and is designed to aid both stock market professionals and everyday investors. It addresses the previously discussed problem statements with the following objectives:

1. To offer an innovative method for making informed stock investments.
2. To accelerate the process of value investment for both stock market experts and novices.
3. To provide short-, medium-, and long-term recommendations with evidence for the stock market for stock market experts and novices.

Our project addresses the issue that investing in the stock market can be daunting for newcomers, who must either acquire substantial domain knowledge to make profitable investments or incur high fees by hiring fund managers. We aim to design a system that facilitates proper investments while significantly reducing both time and financial costs, offering a viable alternative for those new to investing. The system we are developing requires minimal input from the user; they only need to specify the company they are interested in. Our system will then generate a detailed, report-format analysis. This approach reduces the need for hiring fund managers, enrolling in expert-led courses, purchasing investment books, or spending substantial time learning about investments. Our project provides a time-efficient and cost-effective alternative for newcomers to the stock market. The system also indirectly solves the question that newcomers to the stock market may not know where to start learning about investment as there is not a standardized platform and many choices exist in the market. This is because there is no need to study the investment and can also invest in a proper way, which is the motivation of this project. The system also addresses the challenge newcomers face in finding a starting point for learning about investments, given the lack of a standardized platform and the multitude of options available in the market. Users can benefit from our system by observing how analyses are conducted, understanding which factors are emphasized, and more. This educational aspect is a one of the motivation for our project, providing a direct learning pathway for new investors.

Besides that, value investment is a time-consuming task for those stock market professionals and experts. This is because value investment requires in-depth analysis of the companies and markets, which can be time-consuming as the investors need to study, research, and investigate the company in detail. Also, market volatility, which means the market will keep on changing, can make value investing more challenging. The changing market will require investors to conduct continuous research and adjust their trading strategies. Hence, the second objective of the project is to assist experts and normal people in accelerating the process of value investment. As our system automates the data sources collection and analysis. This can generally help the users save time. Even the system gives timely investment recommendations to the users for their references.

Other than that, the system provides short-, medium-, and long-term recommendations with evidence for the stock market for stock market experts and novices. Instead of delivering a prediction of specific numerical values like next-day stock prices or indicators like next-day stock prices going "up" or "down," the system delivers a rating number with more general human-like recommendations for the users, who act as experts to provide the stock suggestions and the evidence to support the suggestions. Hence, the users can take the recommendations and identify whether the supporting information is logical or not before making the investment. This can make decision-making better for experts and regular people alike. One point to note is that the system is designed to provide short-term (3 months to 6 months), medium-term (6 months to 2 years), and long-term (2 years to 5 years) ratings and evidence, respectively. Also, to be noted is that this system is designed to analyze Malaysia's companies.

1.3 Project Scope and Direction

This project utilizes a variety of qualitative data for forecasting purposes, including annual and quarterly reports of companies, relevant news articles, and government policies. To ensure the integrity of our analysis, all data will be thoroughly validated. However, gathering comprehensive information from all companies can be challenging due to time constraints. The scope of this project is specifically outlined as follows:

i. Analyze at least 10 Malaysian companies by incorporating their annual reports from the past five years up to the present.
ii. Collect and review at least one month's worth of recent news related to these companies as additional input for our system.
iii. Include critical government policies such as the "New Industrial Master Plan 2030" and the "New Energy Transition Roadmap," both of which are likely to significantly influence future trends in Malaysia.

Ultimately, this project introduces a system that helps users conduct investment analysis by allowing them to choose from a minimal selection of 10 Malaysian companies. The system provides relevant recommendations backed by solid evidence. It also features customizable analysis options (to be discussed in the following chapter), allowing users to tailor the inputs according to their specific needs. Additionally, the system incorporates a semi-automated data retrieval feature that extracts essential

information such as annual reports, news articles from designated sources, including Bursa Malaysia and The Edge Newspaper. This functionality is particularly useful for administrative staff, enabling efficient data sourcing and minimizing the time spent on manual collection.

1.4 Contributions

Lately, AI has become increasingly popular, partly due to innovations like ChatGPT, which are making AI more integral to our daily lives. A novel concept would be to incorporate ChatGPT into stock market prediction platforms. ChatGPT, as a comprehensive language model, could function similarly to a professional investor, offering recommendations and advice based on analyses such as market trends and sentiments. Its human-like interaction model provides a new user experience compared to traditional algorithms. Therefore, our project utilizes ChatGPT as the main analytics tool for predicting stock market trends in the short, medium, and long term. This can benefit both industry experts and beginners by providing an alternative approach to stock market forecasting. To enhance the user experience, we have simplified the input process to ensure optimal results without the need for users to master specific questioning techniques. Additionally, we offer specialized data handling capabilities, such as annual reports and related company news, that the original ChatGPT does not provide, tailoring our tool to the specific needs of stock market analysis.

From the perspective of investment professionals, our system functions as an assistant that streamlines the analysis process. Experts no longer required to sift through vast amounts of data sources because the tools automatically collect and gather the necessary data to perform recommendations. Hence, it achieved time efficiency for the professionals, allowing them to focus more on investment strategies. Additionally, for beginners in the stock market, our system plays the role of a virtual investment advisor. It offers a novice-friendly pathway for initiating investments while also saving time and money for these new entrants.

Moreover, we demonstrate that ChatGPT can embody distinct perspectives, opening a new avenue for research into how different experts or individuals might interpret the same information differently. We illustrate this by configuring ChatGPT with both

optimistic and pessimistic versions, allowing it to interpret data from these differing viewpoints. For instance, consider the headline, “Tesla Continues to Cut Prices Across Its Car Range.” An optimistic analyst might view this as a sign of Tesla's confidence in its manufacturing efficiency and scaling capabilities, enabling it to reduce prices and potentially increase its market share. In contrast, a pessimistic analyst might interpret the price cuts as an indication that Tesla is struggling with demand issues, necessitating price reductions to boost sales, which could negatively affect the company's profitability and signal deeper challenges.

Other than that, predicting Malaysia's market trends for 2024 through our model represents a significant contribution from our team. In fact, users can gain valuable insights into potential investment opportunities even without specific company details. By analyzing the Malaysia Trends Analysis alone, investors can identify which sectors may be worth investing in. This is largely due to government policies that support certain industries, creating market opportunities and increasing demand within Malaysia. The analysis predicts that the four major trending industries in Malaysia will be Renewable Energy, Electric Vehicles, Semiconductors, and Digital Infrastructure, as illustrated in Figure 1.2 and Figure 1.3.

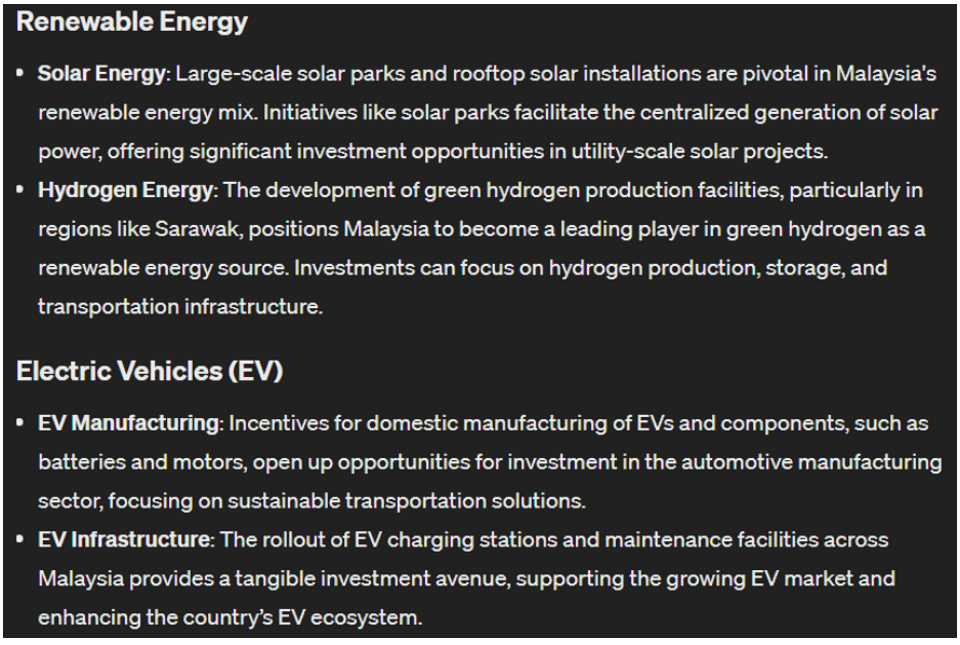


Figure 1.2. First Half of the Predicted Trends in Malaysia 2024.

Figure 1.3. Second Half of the Predicted Trends in Malaysia 2024.

1.5 Report Organization

This report includes a total of five chapters. In Chapter 1, which introduces the project by stating relevant information such as the problem statement and motivation, project scope, project objectives, contribution to readers, and an overview of industry background. Then, a literature review covers investment theories and the recent development of a machine learning system. In Chapter 3, the overall system methodology and design used in this project are discussed. In Chapter 4, the system design, explaining the flows of our system are showed. In Chapter 5, the implementation details of the system are highlighted. In Chapter 6, the system evaluation to evaluate the reliability of the system are highlighted. Last but not least, Chapter 7 summarizes the overall project and following future work.

Chapter 2

Literature Review

2.1 Investment Theories and System Overview

The stock market is the platform for people to buy and sell stocks and shares, as simple as that, but predicting the stock prices is a complex task to implement because countless factors affect the prices, such as company performance, market directions and trends, technical indicators, administrative events from the company, and so on, as explained in the survey paper [1]. Hence, forecasting the stock market is a challenging task because the stock price is not merely influenced by factors within the company but also by public sentiment and the environment. Besides that, the stock market has a theory called the Efficient Market Hypothesis (EMH), as stated in the survey paper [2], which means the stock prices fully reflect all the market information. With this theory, forecasting future stock price movements becomes impossible. However, not all the experts agreed with EMH until today, and the Adaptive Market Hypothesis (AMH) is the theory that is totally contrary to EMH, which means market efficiency will keep changing due to investor behaviours. It also interprets that the market will not strictly follow the "economics principle," but that it can be changed by investor psychology. In short, a lot of aspects can affect the stock price, and they are usually categorized into economic and non-economic factors. Also, some theories from the finance industry prove that it is possible to predict the stock price.

Other than that, we will discuss the logic behind the stock market prediction tool that uses artificial intelligence (AI). We will briefly introduce its concept before delving deeper. The general flowchart of this tool, summarized from survey and review papers [1], [2], [3] employs either machine learning or deep learning systems and is depicted in Figure 2.1. A total of five stages, except the result outputs, will be discussed in the following sections.

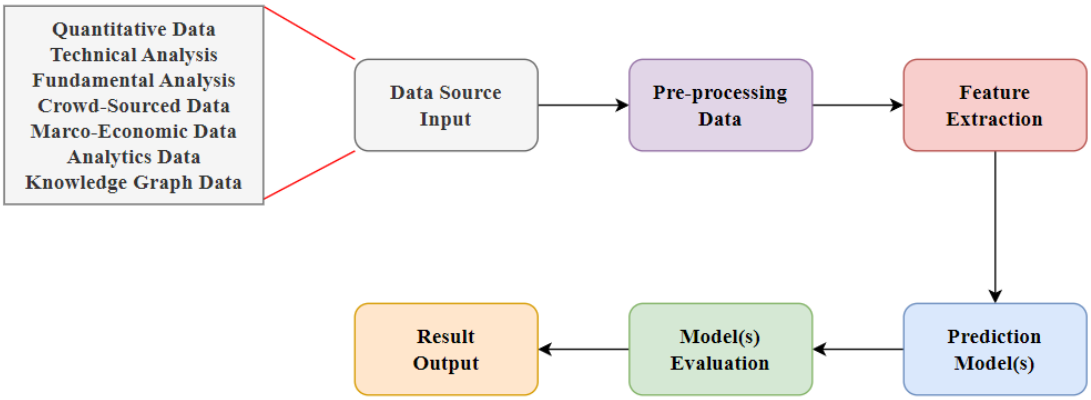


Figure 2.1. Block Diagram of Stock Market Prediction System Overview.

2.2 Data Source Input

Collecting and gathering all the data sources as input for the prediction model is the first and most important step. This is because the different types of input data will have a suitable prediction model to analyse them. Besides that, the type of input data can be divided into seven groups: quantitative data, technical analysis, fundamental analysis, crowd-sourced data, macro-economic data, analytics data, and knowledge graph data as studied from survey and review papers [2], [3]. Each type of data will be explained in detail below.

Firstly, the quantitative data can be historical stock price data, earnings per share, and others, as long as it is numerical information that can be calculated and analysed. These data are used to identify patterns or trends from market history and predict future price movements, which is called back-testing in the industry. Secondly, technical analysis typically involves graphs or indicators that evaluate future price movements, and it had the best performance on short-term predictions. For example, trend indicators, volatility indicators, and so on. Thirdly, the fundamental analysis, as explained before, is often used for long-term investment strategies, and includes company structure, profitability of the company, share index, and others. Fourthly, crowd-sourced data is the data from different opinions of people on the stock market, such as Google Trends, social media posts, news articles, and so on. With this data, it is easier to understand market sentiment, whether positive or negative. Fifthly, macro-economic data are the official statistics that give people information about the economy of a country, such as gross domestic product (GDP), inflation, the consumer price index (CPI), and others. This factor shows the healthy status of the stock market, which

reflects the stability or growth orientation of the stock market environment. Sixthly, analytics data is the data that is analysed on the company, such as recommendations with proofs for buying or selling the stock, and it typically comes from stock market professionals. Seventhly, knowledge graph data is the latest data developed by graph neural networks (GNN) that indicates the relationships between various companies and markets, such as the open sources of FreeBase and Wikidata. One of the use cases is that the investor is able to know how the movement of a stock's price is affected by news.

There are seven categories of data, but not all the data applied to predicting stock prices in the past few years because the data is not easy to obtain from the Internet. In the review paper [3], quantitative data and technical analysis are the most commonly used in the deep learning world, followed by crowd-sourced data. Surprisingly, fundamental analysis is less used in deep learning algorithms, although it is the most common analysis in the financial industry. This is because company reports are rare on the Internet, and the report date is not accurate. However, getting the data could be an issue, but another issue will be that the data is usually too noisy or messy to fit in our prediction models.

2.3 Pre-processing Data

Before handling the noise in the data, information fusion can be used to integrate multiple data sources as input to improve the performance of forecasting tools for stock market prices as compared to the model with only one data input. For example, in the survey paper [2], it proposed a technique to integrate all the quantitative data, fundamental analysis, and technical analysis data together by using an extended Kalman filter (EKF) to predict the next-day stock price trend, and it showed a better performance compared to regression and an artificial neural network (ANN). However, it is hard to set noise covariance matrices that work well in all situations. The noise covariance matrices are applied to describe the noise in the system using statistical properties. Another instance involves merging quantitative data with crowd-sourced information, like Twitter posts, to forecast the stock price trend for the following day. The author used the Granger causality test and vector autoregression (VAR), which increased the accuracy of prediction. The Granger causality test checks if changes in one time series can predict changes in another time series, and the VAR captures the

linear relationships between multiple time series. However, the models needed a total of 2,500 daily tweets in order to be able to make predictions. Hence, various combinations of data as input to the prediction model could have different results. However, identifying the optimal combination of data for model input can be both time-consuming and challenging.

Besides that, the noise of time-series data often exists in the stock market because of the many unreasonable trades that happen there. The review paper [3] provided two techniques to resolve this problem: the wavelet transform and k-Nearest Neighbours (k-NN). The wavelet transform is a signal processing technique that decomposes the stock price time-series data into different frequencies, uses thresholds to remove noise, and reconstructs the signal to a cleaner signal. And the k-NN is using two training datasets but with different labels to compare and contrast to remove noise.

2.4 Feature Extraction

After the data is cleaned, feature extraction also plays an important role because the right feature(s) can directly impact the accuracy of prediction models. In the stock market, feature extraction also means, in the same way as machine learning, finding the key indicators or variables that mostly influence stock prices or market trends, and it is called technical analysis in the market industry. In the review paper [3], the technical indicators, such as the moving average that identifies the direction of market trend and the Relative Strength Index (RSI) that measures the velocity of the price movement changing, are commonly extracted using chart pattern recognition techniques. The chart pattern recognition system detects the pattern from the charts, and it becomes the technical indicator for investors to make predictions about future stock prices.

Besides that, feature fusion is another option to consider instead of removing the feature. Merging the similar features into one feature might improve the accuracy of the prediction models compared to removing one of them. In the survey paper [2], the authors used long-short-term memory (LSTM) and convolutional neural network (CNN)-based models to merge the time-series stock prices with graphical features, which are stock chart images. The LSTM is one type of recurrent neural network (RNN) that can capture the sequence of stock price data, and the CNN can discover important stock chart patterns. With these two models working together, the prediction results

showed improved performance compared to only stock chart images. However, selecting the most important features for model input can be a time-consuming and challenging task as well.

2.5 Prediction Model(s)

After all the data input is well-structured and ready, it is time to put it into the prediction model. Various types of data are best suited for different stock market prediction models. To achieve higher accuracy in the prediction, most of the models studied are combinations. There are a total of 11 models: ANN, CNN, DSS, NN, RNN, SVM or SVR, HMM, filtering classification, fuzzy-based method classification, optimization classification, and NLP.

2.5.1 Artificial Neural Network (ANN)

ANN is the system inspired by the human brain, and its strength is identifying more precisely the relationships between and the factors that affect the stock market and the stock's performance, such as stock prices and dividend yield, compared to many traditional statistical approaches. In the following paragraph, it discusses the previous works by their authors, respectively.

In the survey paper [1], the author, Ticknor, J.L., initially used market trends and technical indicators as input and used the Bayesian regularized ANN to predict the stock prices. But he found out that the performance increased without including the technical indicators. This is the trade-off between technical indicators and the accuracy of the model. Hence, one of the solutions is to add more relevant technical indicators as features to the model, so that it can become more comprehensive and potentially improve its performance. Besides that, Shrivas, A.K., and Sharma, S.K., used the BSE SENSEX data, which is one of the major stock indices in India, and deployed a combination of machine learning algorithms such as support vector machines (SVM), chi-square automatic interaction detectors (CHAID), classification and regression trees (CART), and ANN to predict the stock trends. However, the methods are not considered advanced predictive models because they are not versatile or advanced enough to be used to analyse other stock data, such as Yahoo and BSE 100. Hence, the solution can be including advanced machine learning models such as deep learning can be replaced by some of them to allow the models to handle more complex stock data. Other than

that, Patel, J., et al. introduced a specialized layer called the Trend Deterministic Data Preparation Layer. They used stock trading data and calculated 10 technical parameters to predict the short-term values of stock returns. However, long-term predictions, like quarterly stock predictions, are not suitable for this method. Hence, the extension of the time period for technical parameters to be derived can be considered a solution to the problem of focusing only on short-term predictions. Moreover, Guresen, E., et al. implemented a model called Generalized Autoregressive Conditional Heteroscedasticity (GARCH) to assess the effectiveness of neural networks (NN) in predicting stock market trends. The input features are extracted by a model fusion of hybrid neural networks, the Dynamic Artificial Neural Network (DAN2) and the Multi-Layer Perceptron (MLP). However, the GARCH or Exponential-GARCH (E-GARCH) will fail to predict when they detect the correlated variables or features. Hence, performing a correlation analysis to understand the relationship between the variables could be helpful in identifying the failure factor by GARCH or E-GARCH.

2.5.2 Convolutional Neural Network (CNN)

CNN is a type of ANN, and it is a feed-forward neural network, which means the data flows in only one direction from the input layer through the hidden layer and finally to the output layer. In the following paragraph, it discusses the previous works by their authors, respectively.

In the survey paper [1], the author, Zhou, X, et al., created a generic framework that combines LSTM and CNN to forecast frequent stock market trends. This approach mimics the trader's trading style, and the performance of the model can be improved by using both training and testing datasets. However, it is not versatile enough to handle multiscale data because there are no other predictive models that can handle multiscale conditional data. Hence, integrating a predictive model that can handle multiscale conditions is the straightforward solution to this issue. Besides that, Xu, B., et al. utilized the combination of RNN and CNN as a recurrent convolutional neural network (RCNN) to predict stock market trends. The conversion of categorical to numerical data happens on the first layer of the model. However, this model cannot include financial knowledge to enhance the accuracy of predictions. Hence, one of the possible solutions is to add the technical indicators as features to the models.

2.5.3 Decision Support System (DSS)

It analyses all the factors and shows all the possible outcomes to the users. For example, investors may employ a DSS to guide them in making judicious stock buying or selling decisions. In the following paragraph, it discusses the previous works by their authors, respectively.

In the survey paper [1], the author, Wen, Q, et al., built a smart trading system that combines SVM algorithms with stock box theory. This theory utilized the two-bound forecasts, which means if the stock prices cross the upper bound and lower bound (limits) of the stock box, it could trigger a buy or sell decision for the investors. However, the accuracy of the system is not good enough as it does not have robust estimators. Hence, soft computing techniques such as fuzzy logic and neural networks (NN) can be applied to solve the accuracy problem as they are tolerant of uncertainty in data.

2.5.4 Hidden Markov Model (HMM)

HMM is a statistical model often used to spot patterns. It's really good for tracking time-series data like stock price movement. In the following paragraph, it discusses the previous works by their authors, respectively.

In the survey paper [1], the authors, Gupta, A., and Dhingra, B., introduced a posteriori HMM method to predict stock behaviour based on historical data. The degree of price changes in stock indexes, often called fractional variations, will be used to train the HMM model, and the model will be applied to make maximum posteriori decisions based on stock values. The performance of this model increased when it uses more granular data, such as hour-by-hour and minute-by-minute stock values. However, the correlations between different variables are excluded when building the model. Hence, the consideration of the correlations between different variables can be added to the current model to potentially increase the prediction accuracy.

2.5.5 Neural Network (NN)

NN is a type of machine learning algorithm that is designed to identify patterns within data using a process inspired by the human brain. In the following paragraph, it discusses the previous works by their authors, respectively.

In the survey paper [1], the authors, Chang, P.C., used technical indicators as input to the Evolving Partially Connected Neural Network (EPCNN) to predict stock trends. The historical time series data is learned by the architecture of random connections between neurons that is utilized by EPCNN and also by the evolutionary algorithm that is used to fine-tune EPCNN by selecting the best-performing parameters to improve the performance of the model. However, it cannot handle other time series data as it failed to deploy advanced soft computing techniques. Hence, data preprocessing can solve the time series data issue by adding methods like wavelet analysis and data clustering. Besides that, Chatzis, S.P., introduced a robust prediction system for forecasting different time frames of the stock data. It included multiple machine learning algorithms to do feature extraction and balanced the dataset with bootstrap sampling. However, it cannot process the high-frequency data, which will be updated at short intervals, because the system does not include deep neural networks. Hence, exploration and the addition of deep neural networks to the system can be done to allow the system to process the high-frequency data. Other than that, Shen, W., et al. created a radial basis function neural network (RBFNN) to predict the future stock indices of the Shanghai Stock Exchange. They used the Artificial Fish Swarm Algorithm (AFSA) to optimize the learning process and further optimized it with Genetic Algorithms (GA) and Particle Swarm Optimization (PSO) by fine-tuning the weights in NN. However, the system failed to include qualitative factors into its algorithm, such as market sentiment or news events, to enhance its performance. Hence, the possible solution can be the sentiment analysis, which converts the market sentiment into numerical values as input features and fits into the system algorithm. Moreover, Tsai, C.F., et al. developed a classifier ensemble mechanism that combined multiple classification algorithms, such as homogeneous (especially NNs) and heterogeneous, to predict stock trends. However, this approach is not versatile enough to test other datasets, and non-economic factors are excluded. Hence, technical indicators and non-economic factors can be added to the models to resolve the problem.

2.5.6 Recurrent Neural Network (RNN)

RNN is a type of ANN, and it creates a directed graph that follows the time sequence using the connections between nodes. It can be used in analysing the time-series-based

stock price data. In the following paragraph, it discusses the previous works by their authors, respectively.

In the survey paper [1], the authors, Hsieh, T.J., et al., integrated the RNN with the Artificial Bee Colony (as known as ABC-RNN) algorithm to predict stock prices. There are three major phrases: data preprocessing, where Haar wavelet was used to remove noise from stock prices with time series; feature extraction, where RNN constructed input features using stepwise regression-correlation selection (SRCS); and model optimization, where the weights of RNN were optimized by ABC. However, this system was not able to extract crucial patterns from the data because it lacked advanced pattern selection mechanisms. Hence, a possible solution to the problem is to embed a more advanced pattern selection scheme, such as GA, into the system. Besides that, Chen, W. et al. used the input of various price-related features to the model of RNN with gated recurrent units (GRUs) to predict the stock volatility in the Chinese stock market. However, the model lacked advanced machine learning methods like interdependent latent Dirichlet Allocation (ILDA) for more accurate predictions. Hence, selecting the technical indicators that make the most impact on the stock prices and applying the ILDA to analyse the sentiment using text context are the improvements to this model.

2.5.7 Support Vector Machines & Support Vector Regression (SVM & SVR)

SVM, or SVR, is a supervised learning method with related algorithms for both classification and regression tasks. SVM can classify the prediction outcomes as "buy", "sell", or "hold", whereas SVR can predict the stock market price as long as it is continuous data. In the following paragraph, it discusses the previous works by their authors, respectively.

In the survey paper [1], the authors, Oztekin, combined adaptive neuro-fuzzy inference systems, ANN, and SVM to predict daily stock price movements. However, the stock market data was only from one country and did not test other countries. Hence, the author plans to add more models like decision trees and Bayesian networks for comparison of performance and aims to test the system using the stock market in different countries. Besides that, Ni, L.P., et al. used the fractal selection method to select the important features, handle non-linear patterns in the data, and eventually fit the data into SVM. However, this solution did not consider the micro-economics, such

as the financial status of companies, or the macro-economics, such as GDP. Hence, instead of only including the technical indicators, the sources of micro and macro can be collected as factors in the stock field to enhance the performance of the model.

2.5.8 Filtering

Filtering is the process of removing noise from the data to enhance its quality. Often combining mathematical algorithms or statistical models to analyse past stock prices with technical indicators for predicting future stock price movements. In the following paragraph, it discusses the previous works by their authors, respectively.

In the survey paper [1], the authors, Arévalo, R. et al., utilized fundamental past innovations and flag pattern recognition to design a trading rule for forecasting the stock price. They incorporated a dynamic window scheme for stop-loss and profit updates. They also added an exponential moving average (EMA) indicator for trade filtering. The technical indicators take 15-minute and 1-day timeframes to compute. However, the data snooping problem has been raised and cannot be settled, which means the model overfits the past data and results in poor performance in prediction. Hence, statistical tests like the Reality Check test can be employed to solve the data snooping bias. Besides that, Srinivasan, P., and Ibrahim, P., used the GARCH model to predict the volatility of the SENSEX index over a specific period. The performance metrics indicate that the GARCH model excels at forecasting the SENSEX Index return. However, the common GARCH model has the limitation of heavily relying on historical data and making assumptions such as data normality, which may not always hold true in real-world scenarios.

2.5.9 Fuzzy Logic

Fuzzy logic is used to deal with uncertainty or imprecision. The truth values are not just "true" or "false" (1 or 0). Instead, they can take any real number between 0 and 1, representing degrees of truth. In the stock market, fuzzy logic can be used for volatility modelling as it better adapts to uncertainties and fluctuations in stock prices. In the following paragraph, it discusses the previous works by their authors, respectively.

In the survey paper [1], the authors, Chen, M.Y., and Chen, B.T., introduced the advanced fuzzy time series model to predict stock prices. The model uses granular

computing, entropy-based discretization, and binning-based partition mechanisms to handle the noise data well. However, the model overlooked key factors like technical indicators and financial reports. Hence, the proposed model can be enhanced into a more advanced model by including additional variables like various technical analysis indicators and financial reports. Besides that, Wei, L.Y., et al. designed an adaptive network-based fuzzy inference system (ANFIS) that uses various technical indicators to predict stock price trends. The ANFIS system first selects technical indicators via a correlation matrix, then partitions them using subtractive clustering, and finally optimizes prediction rules through a fuzzy inference system. However, the system was only tested with a specific dataset, so it cannot predict the other stock indices. Hence, the model is able to be tested with other stock markets, such as China and Japan. Other than that, Sadaei, H.J., et al. created an advanced fuzzy set model for trend estimation in fuzzy time series. The data initially turned fuzzy using various fuzzy sets to handle uncertainty, then recognized the relationship between the fuzzy data and eventually defuzzied it for stock market prediction. An imperialist-competitive algorithm was used for training. However, this method cannot remove the seasonality or trends from a time series dataset to make the data stationary. Hence, an extension to the model to include the techniques that can handle the issue, such as the wavelet transform, can be used to analyse the non-stationary data.

2.5.10 Optimization Based

Optimization tends to achieve maximum levels of performance and functionality. Some of the authors utilized this technique in their research on stock market prediction systems. In the following paragraph, it discusses the previous works by their authors, respectively.

In the survey paper [1], the authors, Cheng, C.H., et al., proposed a hybrid forecasting system with a combination of four algorithms: technical indicators based on correlation matrices for predicting stock prices; cumulative probability distribution approach (CPDA) for dividing technical indicators with high similarity; rough set theory (RST) for extracting linguistic rules from technical indicators; and GA for refining the rules from RST to improve stock price prediction. However, this system failed to include other data discretization methods for transforming the continuous data into a discrete format. Hence, the solution is to employ data discretization methods such

as K-means discretization in the data preprocessing stage to resolve the problem. Besides that, Arajo, R.D.A., and Ferreira, T.A., developed the Evolutionary Morphological-Rank-Linear (EMRLF) mechanism for financial time series forecasting and to overcome the random walk theory. The model combines a modified genetic algorithm (MGA) and a morphological rank linear (MRL) filter to minimize time lags. It then uses the Least Mean Square (LMS) algorithm for stock trend prediction. However, this model struggles to regulate anomalies or irregularities based on time series data. Hence, installing an advanced algorithm like LSTM networks could help the model better handle the time-based distortions.

2.5.11 Natural Language Processing (NLP)

NLP is one of the fields of AI that can treat computers as normal people by giving them the skills of comprehensive human language. It can be used for sentiment analysis to analyse stock market sentiment. Inside the NLP, there is a generative pre-trained transformer (GPT) model that is trained on large datasets and generates human-like output for the users. For the latest trends, there are already some popular apps with this technique, such as ChatGPT and Google Bard or Bert, which will be discussed in the following section.

In the survey paper [4], the authors Nguyen et al. performed the experiment of using sentiment analysis on social media to predict the stock market. They used the message board data from Yahoo Finance Message Board to introduce a new feature called "topic sentiment", and historical prices as input, which fits into the Stanford Core NLP for part-of-speech (POS) tagging and lemmatization. However, the system only included a small number of topics and sentiments, leading to a relatively low accuracy rate of 56%. Hence, integrate more features or factors, such as macroeconomic indicators and fundamental analysis, to improve accuracy. Besides that, Bollen et al. applied the Google Profile of Mood States (GPOMS) to capture various moods in Twitter to predict the stock market, and Granger causality analysis was employed to filter out any outlier mood responses from the Twitter users. However, the method did not involve the ground truth for public mood states, which it used to verify or evaluate the accuracy of GPOMS. Hence, this problem can only be solved, in the authors' opinion, by further research that directly assesses public mood states and compares them to those derived from online communities like Twitter.

In the review paper [5], the authors Ko and Chang deployed an LSTM-based sentiment analysis. There are two input data sources: fundamental analysis collected from news articles and PTT forums in Taiwan and technical analysis collected from past stock price data. The sentiment analysis was added to the fundamental analysis by BERT, and then LSTM used the sentiment results together with technical analysis to predict the stock price. However, the emotions in the model are only categorized into positive and negative; there are more types of emotions that can be figured out and categorized. Hence, adding more emotion tags into the model makes the prediction of the model more precise because different types of emotions, such as sadness, fear, and disgust (there are negative emotions), could affect the stock price differently.

2.5.12 Large Language Model (LLM) Review

LLM is one of the categories within NLP, and it is specifically designed to understand human language and generate text like normal human conversation. One of the popular products of LLM is ChatGPT, or to be more specific, GPT-3, GPT-3.5, and GPT-4. ChatGPT is developed by the OpenAI company and acts as an expert to answer all questions that users ask. However, its knowledge has a limit since its training data only includes all the information on the Internet up to September 2021, meaning that the model does not know any incidents that happened after that date. One of the interesting ideas is to apply ChatGPT to the sentiment analysis area and make predictions. The following paragraphs show whether it is better than the existing sentiment analysis tool.

In the article [6], the authors Lopez-Lira and Tang developed a ChatGPT-based sentiment analysis and linear regression to predict stock prices in the short term. The model used the input of Centre for Research in Security Prices (CRSP) daily returns such as stock prices data, trading volumes, market capitalization, and news headlines from RavenPack to perform sentiment analysis, and eventually the linear regression was run to predict future prices, and the results were compared to other existing sentiment analysis methods. As a result, the model achieved a strong correlation between the ChatGPT evaluation and the consecutive daily stock price changes. Furthermore, the authors discovered that ChatGPT sentiment scores outperform other sentiment analysis methods when forecasting stock prices. The proof is in [6, Table 2.1], which indicates the regression analysis from the authors. The ChatGPT (gpt) had

coefficients of 0.118* and 0.134**, indicating significance at the 5% and 1% levels, respectively. The t-statistics are (2.437) and (2.759). It shows statistically significant positive coefficients, indicating a strong and positive relationship between its sentiment scores and next-day stock returns compared to other sentiment analysis methods. However, this is just the beginning of the journey of ChatGPT; a hybrid system can be built using ChatGPT with advanced machine learning tools to enhance the prediction performance.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
gpt	0.118*	0.134**					
	(2.437)	(2.759)					
event_sentiment		−0.065	−0.035				
		(−1.343)	(−0.729)				
gpt_2				−0.014			
				(−0.551)			
gpt_1					0.060*		
					(2.514)		
bert_large						−0.010	
						(−0.196)	
bert							−0.147*
							(−2.487)
Num.Obs.	46 094	46 094	46 094	46 094	46 094	46 094	46 094
R2	0.220	0.220	0.220	0.220	0.220	0.220	0.220
R2 Adj.	0.160	0.160	0.160	0.160	0.160	0.160	0.160
R2 Within	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R2 Within Adj.	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AIC	263 554.6	263 554.3	263 561.8	263 562.3	263 557.6	263 562.5	263 555.1
BIC	292 260.4	292 268.8	292 267.6	292 268.0	292 263.3	292 268.3	292 260.8
RMSE	3.93	3.93	3.93	3.93	3.93	3.93	3.93
Std.Errors	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno	by: date & permno
FE: date	X	X	X	X	X	X	X
FE: permno	X	X	X	X	X	X	X

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.1. Regression of Next Day Returns on Prediction Score. Source: Adapted from [6]

In the article [7], the authors Steinert & Altmann used GPT-4 and BERT to do sentiment analysis based on microblogging messages such as those on the website Stocktwits for predicting Apple and Tesla's daily stock price movements for the year 2017. The messages were undergoing different methods of handling according to their respective models. For GPT-4, the model removes all URLs, references to websites, duplicate entries in tweets, and images. All text is converted to lowercase. And for BERT, additional cleaning steps are taken, such as the removal of hashtags, mentions, and special characters. Then prompt engineering was utilized to unlock the full power of LLM models. Prompt engineering is used to improve the performance of LLM by refining the input prompts. After the outputs of LLM, logistic regression is used to evaluate whether the outputs have a significant impact on the stock price movements. The outcome is the movement of stock prices, categorized as either "up" or "down". And the paper highlighted that GPT-4 is performing well compared to BERT because GPT-4 is able to understand and capture the nuanced sentiments from the text to provide

a more precise answer. It is also shown that the GPT-4 is able to understand the terminology in finance, probably due to their training set being different from others. However, considerations about deploying GPT-4 in projects can be made before applying because the costs are generally not cheap.

2.6 Model(s) Evaluation

The purpose of the model evaluation is to measure or assess the performance of the model to identify its effectiveness and prediction accuracy. In the review paper [3], it was categorized into some types. First, classification metrics are used to measure the effectiveness of classification problems. For example, their metrics are accuracy, precision, recall, F1 score, and others. In stock market prediction, accuracy can be applied, which counts the number of correct predictions for the directional change in stock price movement. Secondly, regression metrics are used to measure the effectiveness of continuous numerical value prediction. For example, their metrics are mean absolute error (MAE), mean squared error (MSE), mean absolute percentage error (MAPE), and others. This is the most common approach to evaluating the performance model in stock market prediction. Thirdly, profit analysis is used to measure the effectiveness of trading strategies based on predictions. It used two aspects to do the analysis: return and risk. It calculates how much money a strategy could make as a percentage and the likelihood that the strategy will result in a loss. One of the financial metrics used to do profit analysis is the Sharp Ratio.

In the survey paper [1], each of the systems is evaluated with suitable metrics to measure its performance. However, we discussed some of it in the following: A merging of fundamental past innovations and flag pattern recognition to design a trading rule for forecasting the stock price, which was discussed before in the filtering section. The system used accuracy to measure, and it achieved a range of 90% to 100% under optimal conditions. Besides that, there is an EMRLF mechanism to do prediction, which was discussed before in the optimization section. The system used MAPE to measure, and it achieved a range of below 0.5, which means it has small errors between predicted and actual values. As a result, it is a signal of more accurate prediction. Other than that, an ABC-RNN algorithm to predict stock prices, which was discussed before in the RNN section, The system used root mean square error (RMSE) to measure, and

it achieved a range of below 0.3, which indicates the same meaning of MAPE: the smaller the values, the greater the accuracy of the system.

In the survey paper [4], two methods for predicting stock prices were discussed within the NLP section. The Stanford Core NLP method was evaluated based on accuracy and achieved a rate of 56%. On the other hand, the GPOMS method was assessed using both accuracy and MAPE as metrics. It attained an accuracy of 86.7% and also reduced the MAPE by over 6%.

In the review paper [5], LSTM-based sentiment analysis with the BERT system was discussed under the NLP section. It was evaluated based on RMSE and achieved a 12.05% average accuracy improvement compared to Long Short-Term Memory with Peephole Connections (LSTMP)-based sentiment analysis.

In the article [7], the accuracy comparison between GPT-4 and BERT was made using the accuracy metric. In Table 2.2, it also compares the benchmark model using simple buy-and-hold strategies, which are called "naive" in the table. The p-values can also be calculated using the Naive as a benchmark. "n_{test}" is the number of observations in the test samples. GPT-4 performs better than both BERT and the Naive strategy, with an accuracy range of 66.27% to 71.47%. Low p-values indicate that GPT-4's performance is statistically better than the Naive strategy.

Begin test	Naive	BERT	$p_{Bert\|Naive}$	GPT-4	$p_{GPT\|Naive}$	n_{test}
April	51.85%	65.87%	2.29×10^{-4}	70.11%	3.85×10^{-5}	378
May	52.06%	66.47%	9.10×10^{-3}	71.47%	4.91×10^{-5}	340
June	51.69%	65.54%	9.48×10^{-2}	69.26%	1.42×10^{-2}	296
July	51.59%	65.87%	1.91×10^{-4}	66.27%	1.06×10^{-3}	252
August	50.47%	66.51%	4.87×10^{-2}	66.51%	8.58×10^{-4}	212
September	46.99%	65.66%	5.23×10^{-2}	66.87%	6.02×10^{-4}	166

Table 2.2. Comparison of accuracies for stock prices movements. Source: Adapted from [7]

Chapter 3

System Methodology/Approach

3.1 System Design Diagram

3.1.1 System Architecture Diagram

CRISP-DM methodology (Cross-Industry Standard Process for Data Mining), which includes phases such as understanding the problem, data understanding, data preparation, modelling, evaluation, and deployment. Our system's approach adopts some elements of the CRISP-DM framework, encompassing stages like the collection of data sources to comprehend data, data pre-processing for preparation, the construction of the model, its subsequent evaluation, and the use of a GUI that aligns with the deployment phase of CRISP-DM's standard methodology.

The general work procedures of the system in the Figure 3.1 consist of two segments: admin and user. The admin section, highlighted in light red, is dedicated to tasks such as acquiring and preprocessing data, managing databases, and refining the model for optimal results, with a focus on assessing and improving model performance. Conversely, the user area, indicated in light blue, solely involves interaction with the system through a GUI, enabling users to satisfy their curiosity and gather information about different company stocks.

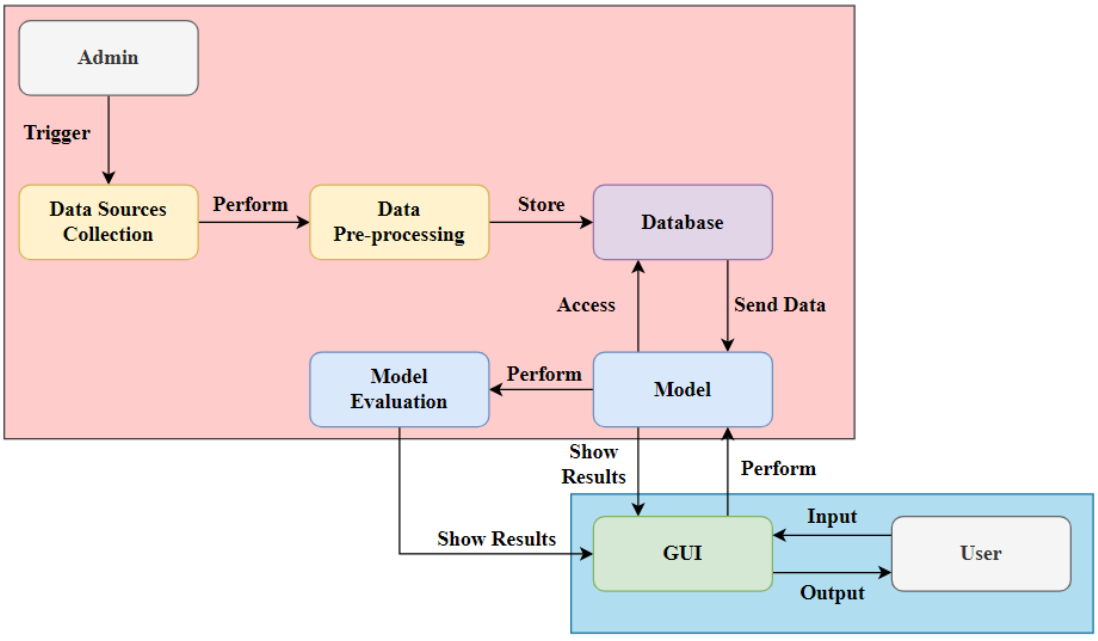


Figure 3.1. General Work Procedures of System.

3.1.2 Use Case Diagram and Description

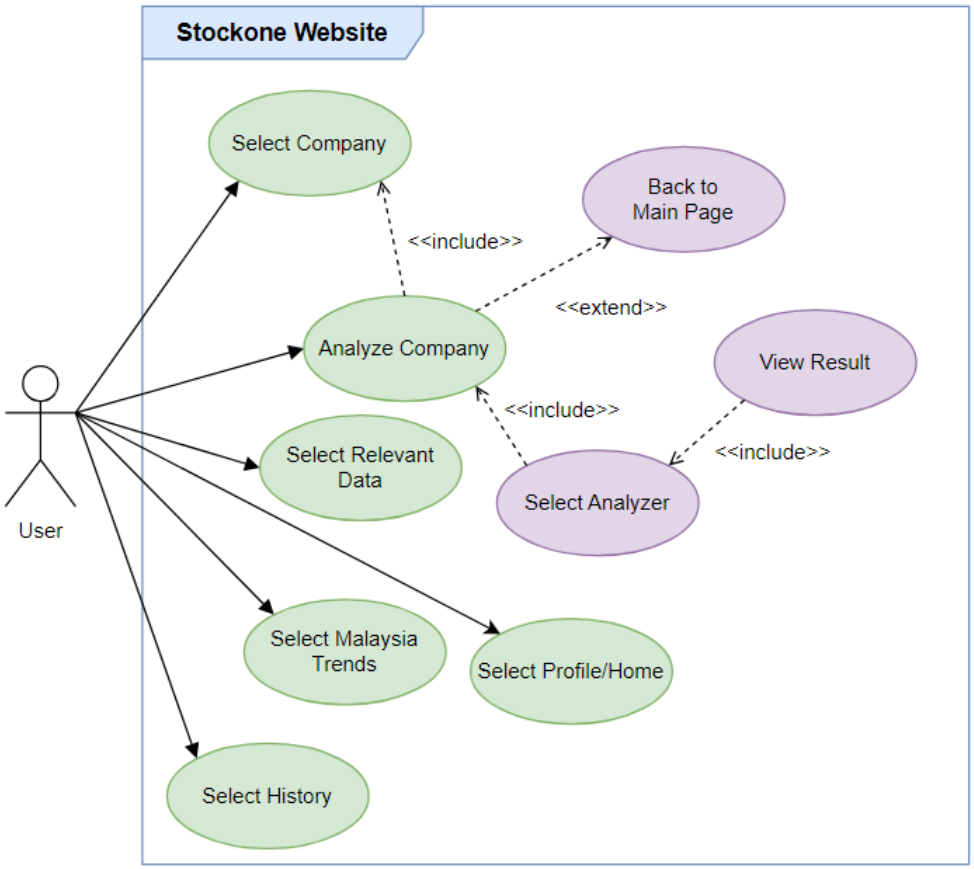


Figure 3.2. Use Case Diagram for the Website.

The main users of the system interface on our website can be either professionals or investors. Users of the system can choose a company to analyze; note that the “Analyze Company” feature activates only after a company is selected. Once the company is analyzed, the user can choose between two types of analysis perspectives: optimistic or pessimistic. The selected analysis type will determine the results displayed. If no analysis type is chosen, users can return to the main page of the selected company. Additionally, users can access relevant preprocessed data for each company, including summarized annual reports, quarterly reports, credibility assessments, competitor analyses, and related news. Furthermore, users have the option to explore predicted Malaysian trends for 2024 and view their historical queries within the system. This comprehensive functionality ensures a user-friendly and informative experience. Additionally, if users select the "Profile/Home" option, they will be redirected to the homepage.

3.1.3 Activity Diagram

Main Function as Analyze the Company

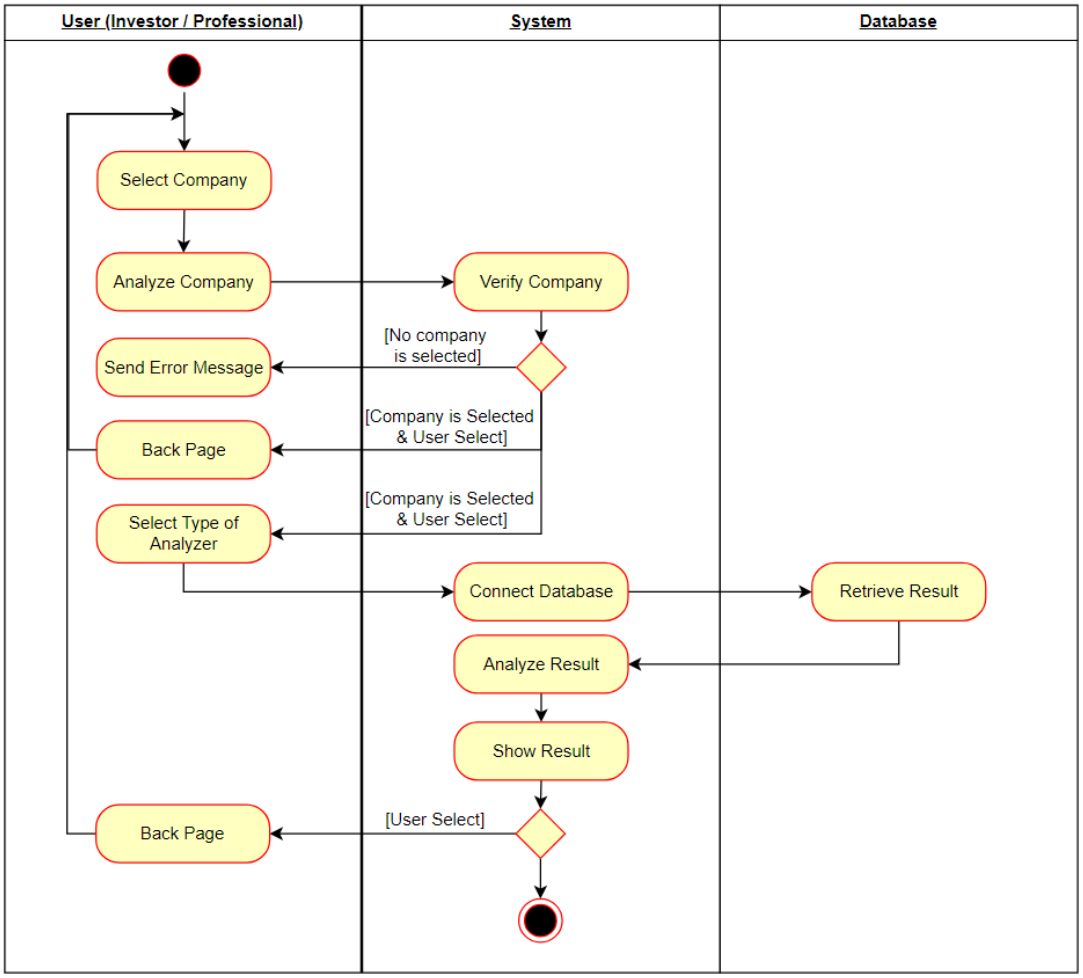


Figure 3.3. Activity Diagram of Analyzing the Company.

Side Functions (Relevant Data, Malaysia Trends, History)

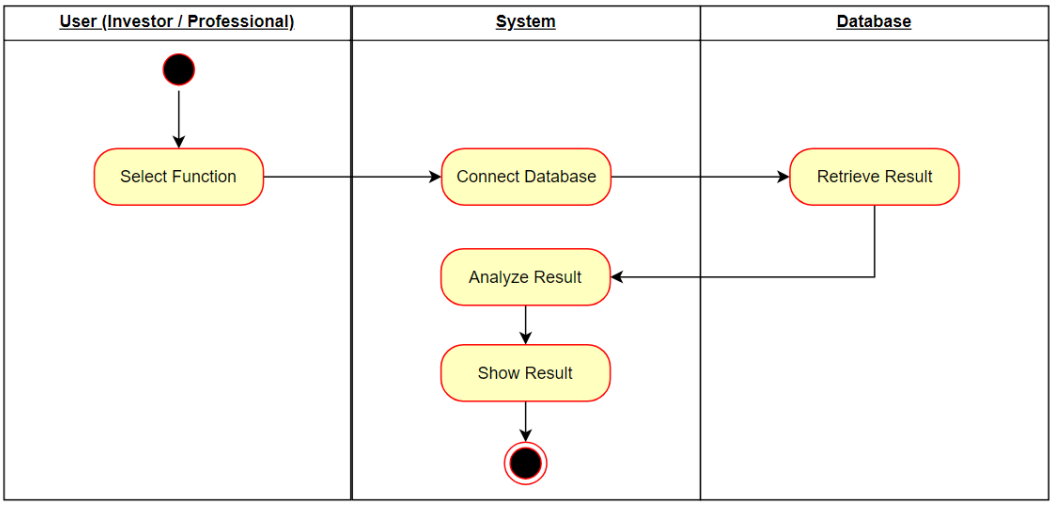


Figure 3.4. Activity Diagram of Side Function.

Chapter 4

System Design

4.1 System Block Diagram

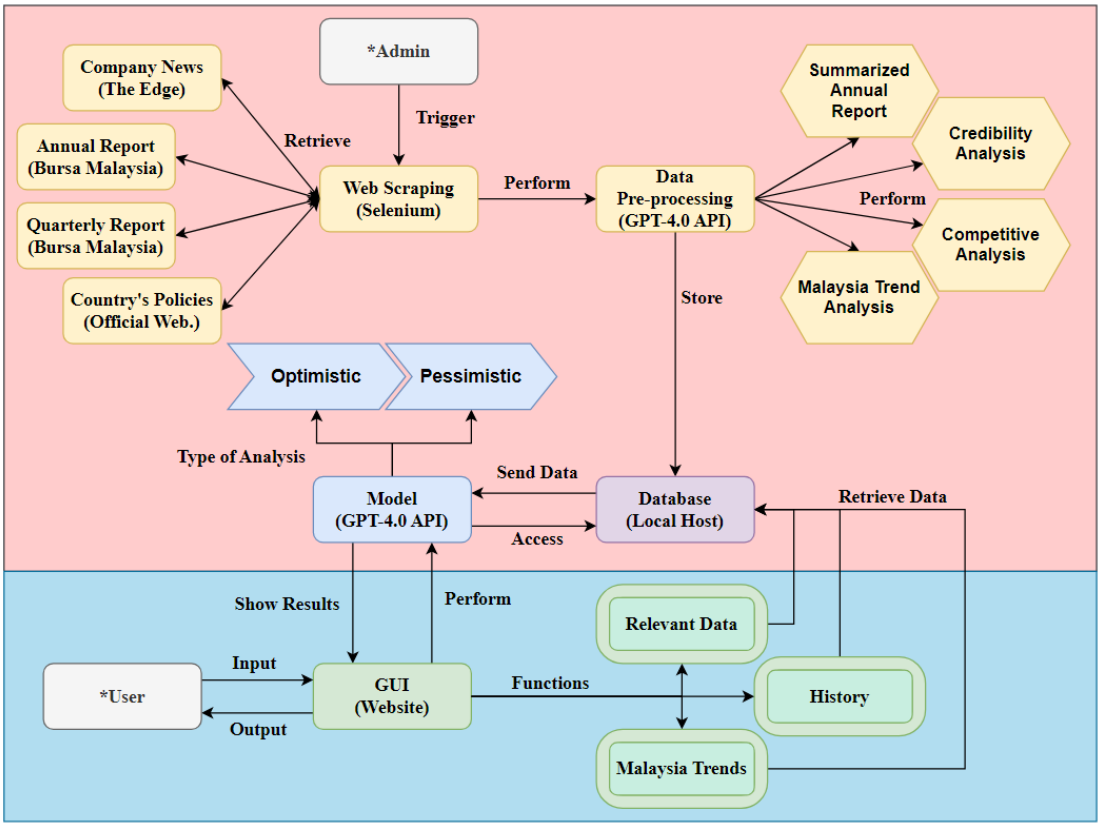


Figure 4.1. Block Diagram of System.

4.1.1 Web Scraping (Selenium)

The purpose of the web scraping tool is to automatically collect and gather information from the Internet. There are four sources of information that the system will get: company news, the annual report, quarterly report, and the country's policies. For company news, we utilize "The Edge Malaysia," a well-known Malaysian online platform specializing in business and financial news. Annual reports and quarterly reports are sourced from "Bursa Malaysia," the primary investment platform in Malaysia, where listed companies are mandated to submit their annual reports within four months after their financial year ends, making it a vital resource for our data collection. In addition, we gather national policy information from the official websites of major governmental departments. We focus on four key departments: Transport,

Education, Science, Technology, and Innovation, and Economy. From these sources, we also collect information on mainstream policies such as Budget 2024, the Madani Economy, the New Industrial Master Plan 2030, the Twelfth Malaysia Plan (RMK12), and the New Energy Transition Roadmap.

4.1.2 Data Preprocessing (GPT-4.0 API)

Instead of completing all tasks on the "Model" depicted in Figure 4.1, we have opted to distribute the tasks across multiple GPTs to preprocess the data before it serves as input for analysis. This approach not only reduces the number of input tokens required but also lowers the costs associated with calling the GPTs. We have designated four separate GPTs to handle distinct tasks, as follows:

i. Summarized Annual Report

In the company annual report, we concentrate on three principal sections: the board of directors, the chairman's statements, and the management discussion & analysis. For the board of directors, the designated GPT is tasked with retrieving their names and respective roles. In the chairman's statements, the GPT focuses on extracting insights about the company's performance, strategic direction, governance practices, challenges and opportunities, future prospects and growth strategies, sustainability, corporate social responsibility (CSR) and ethical practices, risk management, and industry trends and competitive landscape. The management discussion & analysis covers similar areas as the chairman's statement but also includes detailed evaluations of their business segments and annual performance. After processing, the results should be organized into three distinct sections: Board of Directors, Chairman’s Statement, and Management Discussion & Analysis. This structured format will facilitate efficient storage in our database.

ii. Credibility Analysis

Credibility analysis is conducted to verify whether the management or chairman of a company has met their stated targets. For example, if the chairman commits to digitalizing operations but no digitalization efforts are evident in the following year, this will reflect poorly on their credibility. To perform this analysis effectively, it is essential to review multiple years of annual reports to check the consistency of their

stated goals against actual achievements. In our system, the input consists of multiple years of annual reports. The GPT is tasked with extracting forward-looking statements, challenges faced, any subsequent follow-through, and the implementation of future directions. The output will provide an analysis of the implementation of future directions, interpretation of sentiments, and the challenges and solutions identified. It's important to note that this analysis does not include numerical evaluations of the company's credibility, focusing instead on qualitative assessments.

iii. Competitive Analysis

The purpose of conducting competitive analysis is to enhance the realism of our results. Focusing solely on the selected company might not provide sufficient insight into its industry positioning. By comparing strengths and weaknesses relative to its competitors, users gain a clearer understanding of the company's competitive landscape. However, in our system, competitors are manually identified, and typically only one or two are compared to. We utilize the latest quarterly reports to assess current industry performance, adapt credibility analysis to evaluate the integrity of management, and review recent news to compare their current activities. The GPT is tasked with comparing the selected company and its competitors by assessing their market position, performance, financial status, strategic initiatives, credibility, and risk. The output will be a performance ranking within the industry, ranging from 1 (highest) to 2 or 3 (lowest). Each ranking will be accompanied by a detailed justification, providing clear insights based on the comparative analysis.

iv. Malaysia Trend Analysis

Instead of using the GPT-4 API, we employ ChatGPT-4 to analyze market trends in Malaysia based on government-published policies, leveraging its capability to retain extended context, a feature not available in the API version. However, we encounter challenges due to the broad scope of these policies, which may lead ChatGPT-4 to randomly select from a wide range of targeted industries for prediction. For instance, the "New Industrial Master Plan 2030" covers various sectors like aerospace, food processing, and automotive. To enhance our analysis, we incorporate "progress news" to provide current context. For example, last year, Intel announced the construction of its largest 3D chip packaging facility in Penang, Malaysia, which is relevant to the

semiconductor sector included in the master plan. This integration of policy analysis with recent developments enables the model to identify and predict market trends more accurately. The results are shown in Chapter 1's contribution.

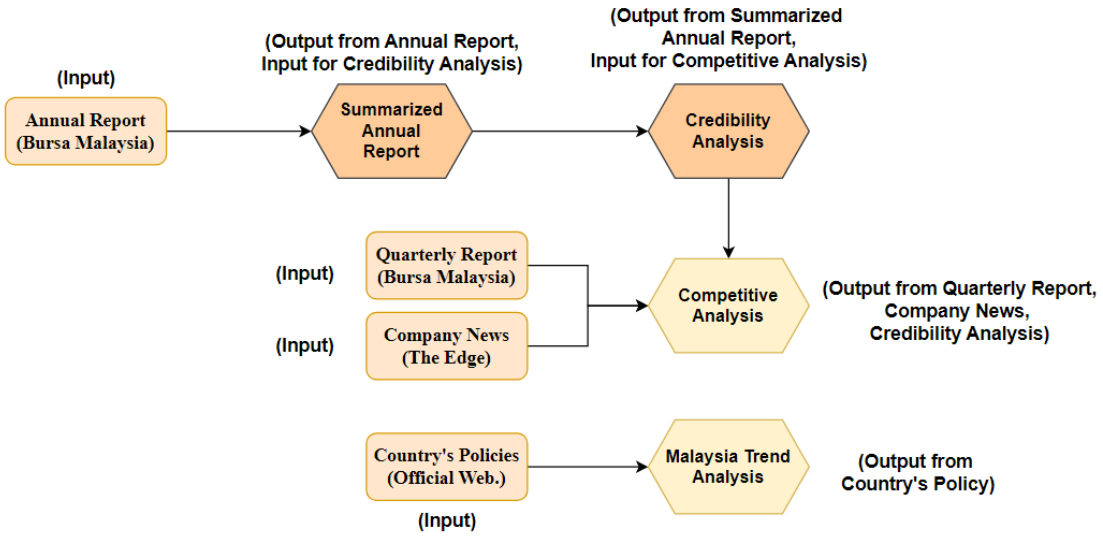


Figure 4.2. Summarization of Input for Each Preprocessing Analysis

4.1.3 Database (PostgreSQL – Local Host)

The database will be hosted locally, utilizing PostgreSQL as the primary relational database management system. The database structure is organized into eight tables: "Company," "Annual Report," "News," "Country News," "Credit Company," "Industry Competitor," "Quarter Financial Report," and "History." The "Company" table serves as the parent table and includes essential details such as stock code, company name, sector, and the latest report year to prevent redundant report downloads. The remaining tables except the "Country News," which are child tables, store detailed information pertaining to annual reports, news articles, national trends, credibility analyses, industry competitors, quarterly financial reports, and user history, respectively. These child tables maintain a one-to-many relationship with the "Company" table, accommodating multiple entries per company. A diagram of this structure is provided in Figure 4.3 for clearer visualization.

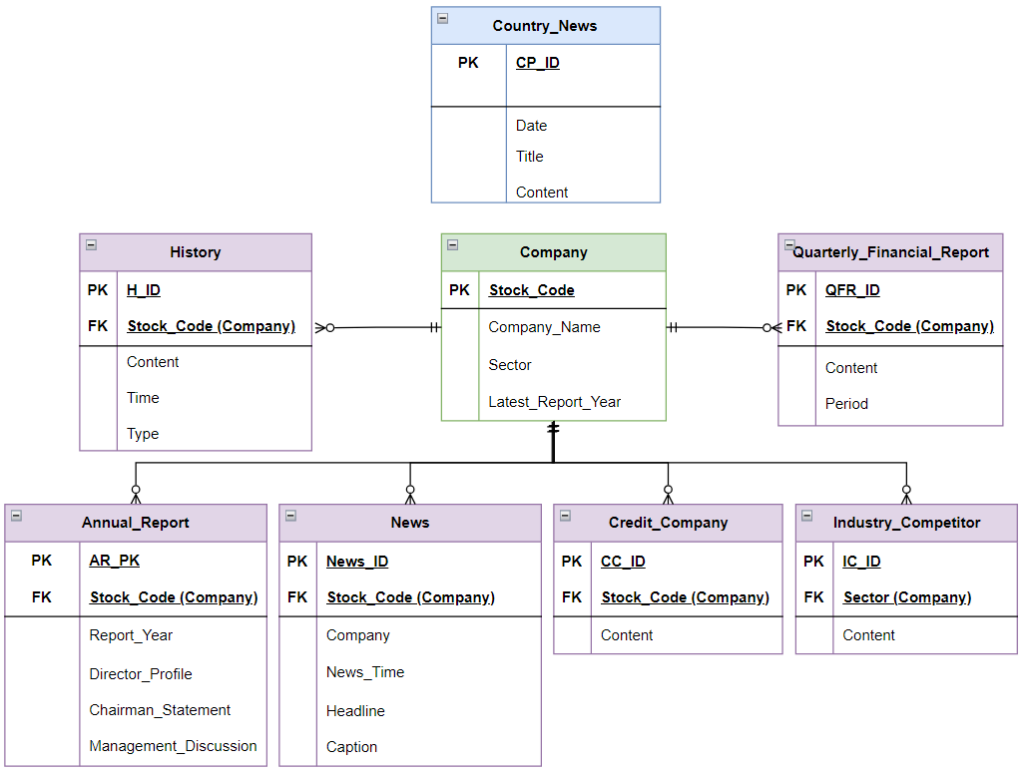


Figure 4.3. Entity Relationship (ER) Diagram of System Database.

4.1.4 Model (GPT-4.0 Turbo API)

GPT API will be our main model in the system, which is used to make analyses on data and give recommendations and supportive evidence to the users. There are also other LLM models, such as BERT, but ChatGPT will be our choice. This is because the results of accuracy proved that ChatGPT will be better than BERT, as shown in Tables 2.1 and 2.2. However, the cost will be an issue because it costs $0.01 per 1000 tokens input and $0.03 per 1000 tokens output for the GPT-4 Turbo API model. This is costly compared to others and might need careful pricing when deploying the model.

Additionally, the model is designed to operate like a skilled investment analyst, primarily processing qualitative data from various sources such as the latest annual reports, the most recent quarterly reports, credibility analyses, competitor analysis, and trends in Malaysia. It leverages its database, with the current knowledge cut-off for the GPT-4 Turbo Model being December 2023. The model is programmed to carefully consider factors such as leadership changes, market position, and management credibility while focusing on generating clear and actionable investment recommendations. The model's outputs are categorized into three-time frames: short-

term (3-6 months), medium-term (6 months to 2 years), and long-term (2-5 years). Each period concentrates on distinct aspects such as recent financials for the short term, market positioning and credibility analysis for the medium term, and strategic initiatives and business alignment with the trends for the long term. The system issues recommendation ratings on a scale from 0 to 10, accompanied by supporting evidence, and includes an assessment of the company's credibility.

Moreover, we have developed two versions of the analyzer: optimistic and pessimistic. The optimistic analyzer emphasizes the positive aspects of information, such as financial growth, alignment with market trends potentially leading to increased demand, and a generally positive long-term outlook. In contrast, the pessimistic analyzer focuses on identifying vulnerabilities within the management, exploring potential risks to the company, and critically assessing whether the company is truly prepared to scale production to meet increasing market demands, despite alignment with current trends. These two approaches are distinctly designed to provide contrasting perspectives on the same data.

4.1.5 Model Evaluation

In our model evaluation method, we assess the qualitative data provided by our system, specifically focusing on the recommendations and justifications it delivers to users. This is to ensure that the information relayed by the system is truthful and entirely based on facts. In subsequent sections, we conducted tests on six companies using both the optimistic and pessimistic versions of our analyzer.

4.1.6 GUI (Website)

Ultimately, users can inquire about a company's stock market performance using the graphical user interface. This will yield outputs such as recommendation ratings with supporting evidence and an overall credibility assessment of the company. Additionally, the system offers users three selectable functions: Relevant Data, Malaysia Trends, and History. The Relevant Data function displays all information stored in our database, including summarized annual reports, related news, competitive analysis, and more. The Malaysia Trends function provides predictions for Malaysia's market trends in 2024. Meanwhile, the History function allows users to access and review their past analyses.

CHAPTER 5

System Implementation

5.1 Hardware Setup

The hardware involved in this project is computer only.

Description	Specifications
Model	Lenovo IdeaPad L340-151RH Gaming
Processor	Intel Core i5-9300HF
Operating System	Windows 11
Graphic	NVIDIA GeForce GTX 1050
Memory	8GB DDR4 RAM
Storage	256GB SSD

Table 5.1. Specifications of Laptop.

5.2 Software Setup

The core software involved in this project with brief explanation are shown below.

1. Selenium Library

In this project, Selenium, a freely open-source library, is employed to automate the process of gathering data sources. This tool facilitates the scripting of web navigation and interactions with elements like forms and buttons on web pages, streamlining the data extraction process.

2. PyPDF2 Library

PyPDF2 is a Python library that is used to manipulate PDF files in a coding environment. It allows for tasks such as reading data from PDFs and splitting and merging PDF documents. In this project, we use this library to extract the specific information that is desired from the annual report of the company in PDF files.

3. Psycopg2 Library

Psycopg2 is a PostgreSQL database adapter for the Python programming language. It is a popular choice for connecting Python applications to a PostgreSQL

database. Psycopg2 allows developers to execute PostgreSQL commands from Python, enabling tasks such as querying the database, inserting or updating data, and handling transactions. In this project, we use this library to communicate with our PostgreSQL cloud database. Storing the structured data that is extracted from the annual reports or company news.

4. Visual Studio Code

Visual Studio Code, created by Microsoft, is a complimentary coding editor supporting various languages including C, C++, Python, among others. In our project, we use this software to implement Python coding for web scraping, PDF data retrieval, communicating with cloud databases, and also implementing the GPT model.

5. pgAdmin 4 (Local Host PostgreSQL Database)

pgAdmin 4 is a comprehensive database management tool designed for PostgreSQL, which is one of the most advanced open-source database systems. It provides a user-friendly platform where users can execute SQL queries, manage database objects, and visualize data structures. pgAdmin 4 supports various PostgreSQL features and is compatible with multiple operating systems, making it a versatile tool for database administrators and developers alike.

6. OpenAI

OpenAI develops the most advanced AI models and technologies; one of their famous products is ChatGPT, which caused a significant impact in various industries for its sophisticated language processing capabilities. The OpenAI API service serves as our main model for analyzing documents and information to generate recommendations and evidence for users.

7 . Node.js

Node.js is an open-source, cross-platform runtime environment for executing JavaScript code outside of a browser. It is equipped with a large ecosystem of libraries via the npm (Node.js package manager), enhancing its functionality for various development tasks. This makes Node.js a popular choice among developers for creating web servers, networking tools, and more. Our team uses Node.js to host the local server.

Particulars	Tools
Operating System	Microsoft Window 11
Integrated Development Environment (IDE)	Visual Studio Code
Programming Language	Python
Libraries	Selenium, PyPDF2, Psycopg2, OpenAI, OS, Re, JSON, Time, Datetime, Shutil, Glob
Database	pgAdmin (PostgreSQL)
User Interface	HTML5, CSS3, JavaScript

Table 5.2. Software Tools for Development.

5.3 Setting and Configuration

When using the GPT API, one important parameter is called "temperature," which controls the randomness or creativity of the model's responses. This parameter is set as a numerical value that typically ranges from 0 to 1. A low temperature, closer to 0, results in more deterministic, consistent, and predictable responses from the model. This setting is ideal for tasks requiring high precision where less variability is desired. On the other hand, a high temperature, closer to 1, makes the model's responses more varied, random, and creative. This higher setting allows the model to produce a broader array of outputs and engage in more experimental language generation, which is beneficial for creative tasks or scenarios where exploring diverse response options is preferred.

In our system, we assign different temperature settings to different GPTs based on the tasks they perform. For instance, for all data preprocessing tasks, the temperature is set to 0. This ensures that the responses are strictly factual and reliable, avoiding any generation of inaccurate or fabricated content. For interactive sessions where users consult the analyzer, we set the temperature to 0.2. This allows for responses that are primarily based on factual justifications, while still permitting a slight degree of creative interpretation in the answers provided.

5.4 System Operation (with Screenshots)

Develop a sleek and user-friendly website interface, as depicted in Figure 5.1, for Stockone, our specialized qualitative data analysis system.

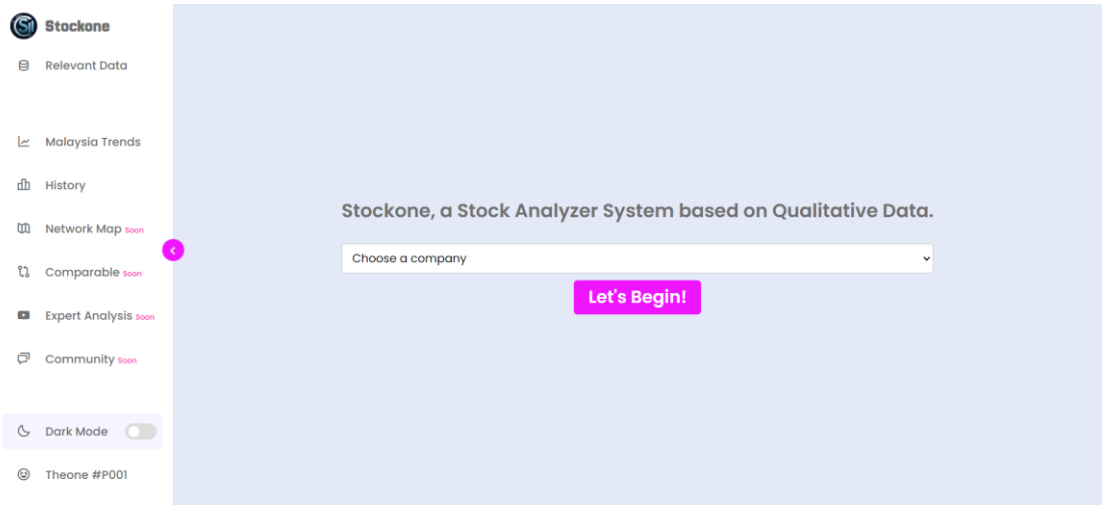


Figure 5.1. Home Page for the website.

Users can begin by selecting a company from our database using the search bar. As depicted in Figure 5.2, there are 11 companies available for selection, some of which are categorized into different industry sectors such as renewable energy, manufacturing, consumer products, and utilities. After choosing a company, users can click the “Let’s Begin!” button to proceed. Next, they must choose the type of analysis—Optimistic or Pessimistic. Each type is accompanied by an image and a description that explains how the input data will be interpreted, enhancing the user’s understanding of the analytical process. Additionally, users have the option to return to the homepage by clicking the “Go Back” button, as illustrated in Figure 5.3.

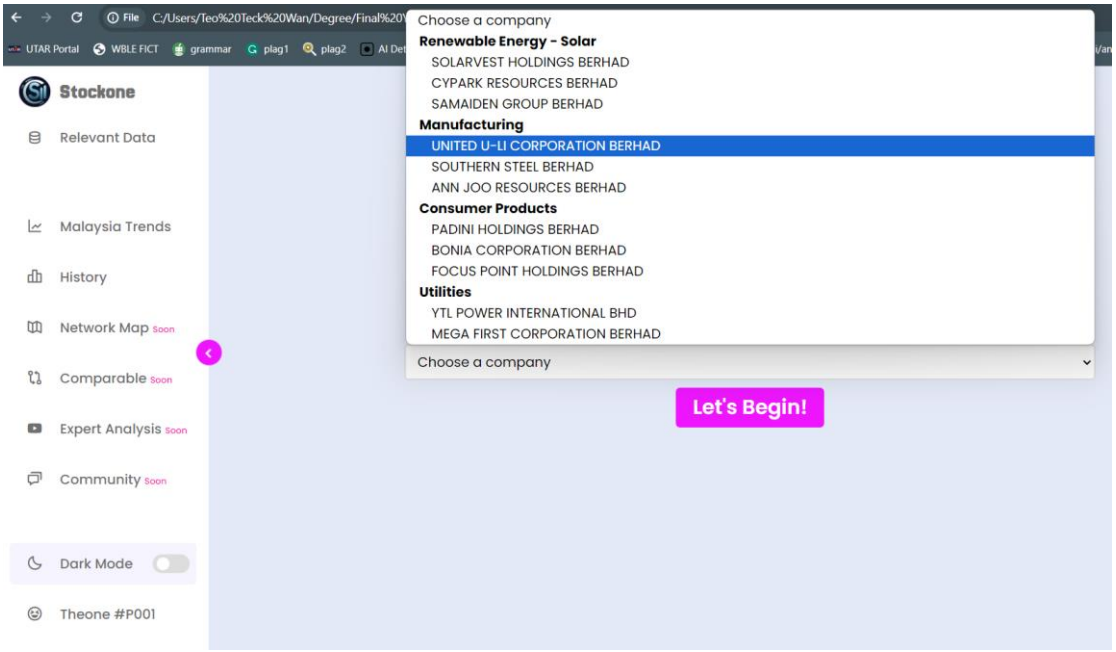


Figure 5.2. Selecting the company for analysis.

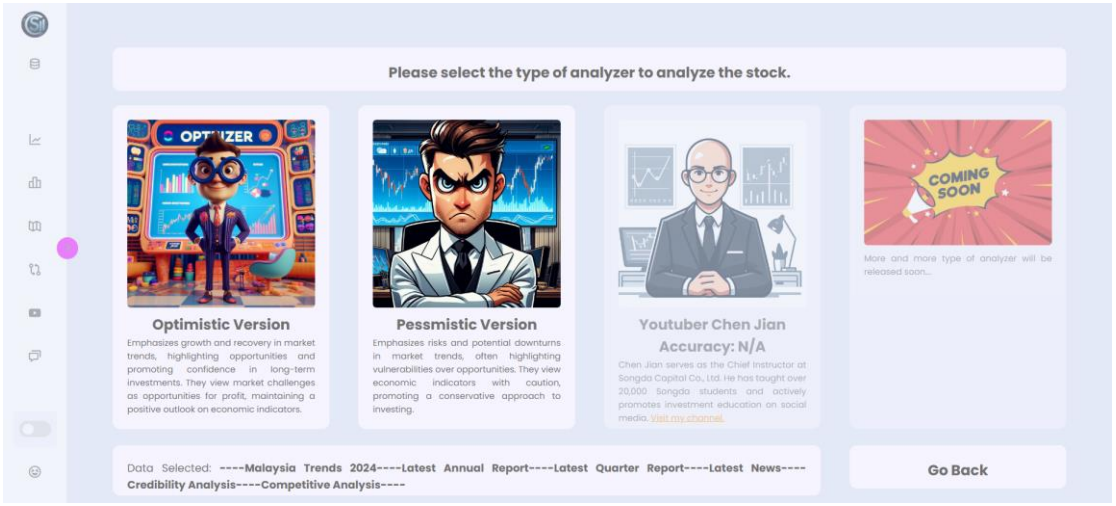


Figure 5.3. UI of selecting the type of analyzer.

After selecting the type of analysis—for instance, opting for the optimistic analysis for Cypark—the system will display the results, which typically take about 30 to 50 seconds to appear, depending on server response times, as shown in Figure 5.4. Additionally, users can return to the homepage at any time by clicking the home or profile icon located at the bottom-left of the screen.

Figure 5.4. Results of Optimistic Analysis.

Other than that, we also provided the light or dark mode for user to switch according to their preferences as shown in Figure 5.5.

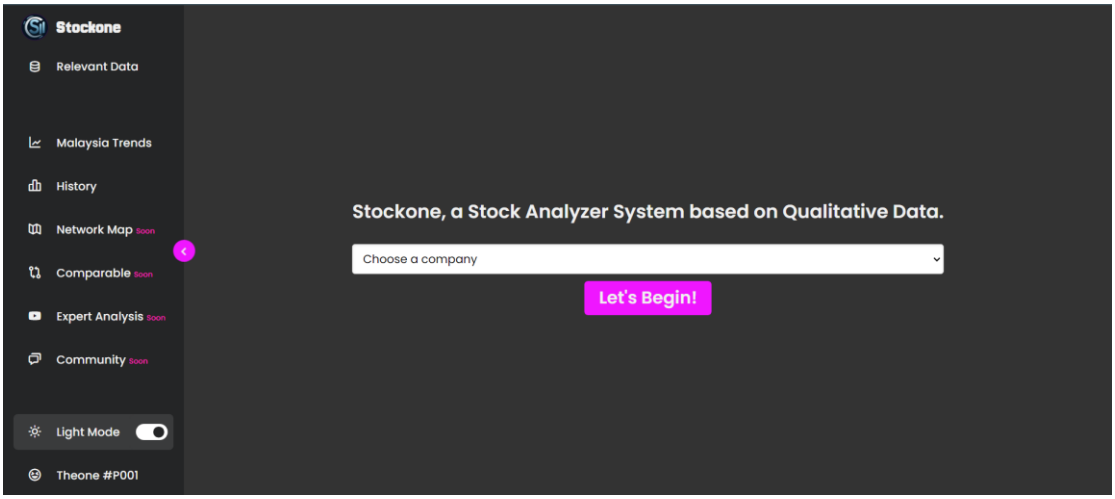


Figure 5.5. Dark mode interface of website.

The "Relevant Data" function in the sidebar allows users to access various types of information—such as annual reports, quarterly reports, related news, credibility analysis, and competitor analysis—while waiting for the analysis results, as illustrated in Figure 5.6. Additionally, the sidebar change and serves as a filter, enabling users to specify and view only the information they require.

Figure 5.6. Function: Relevant Data

Other than that, the Malaysia Trends function shows the predicted trends for 2024 in Malaysia, as shown in Figure 5.7.

Figure 5.7. Function: Malaysia Trends

Moreover, the "History" function displays past records and queries made by the user. Other features in the system are labeled as 'soon,' indicating that they are not currently available but are planned for future implementation. This is clearly depicted in Figure 5.8, highlighting areas of potential expansion for the platform.

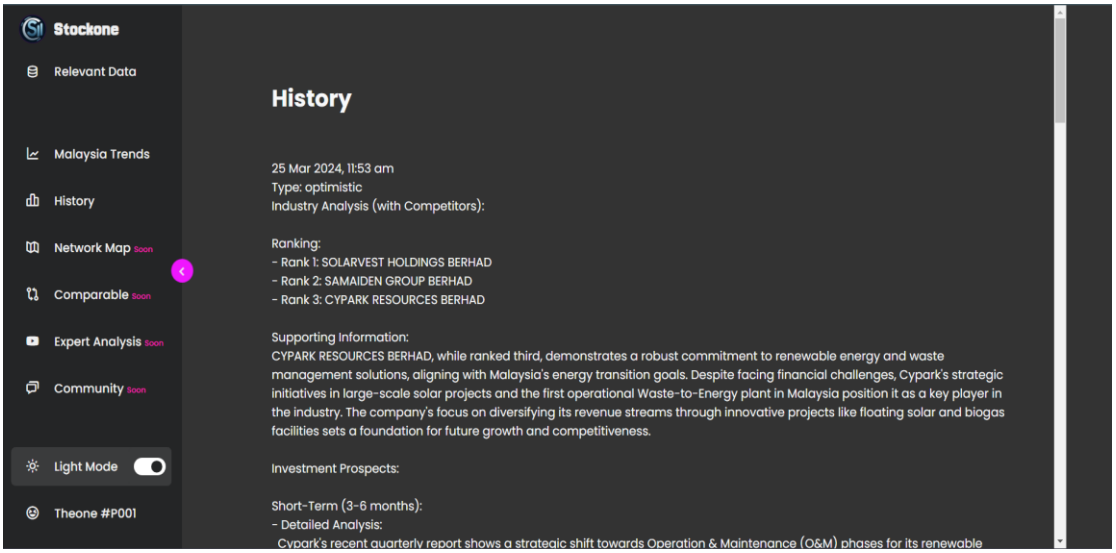


Figure 5.8. Function: History

Figure 5.9. Sample Output from System (Cypark with Optimistic) – Part 1

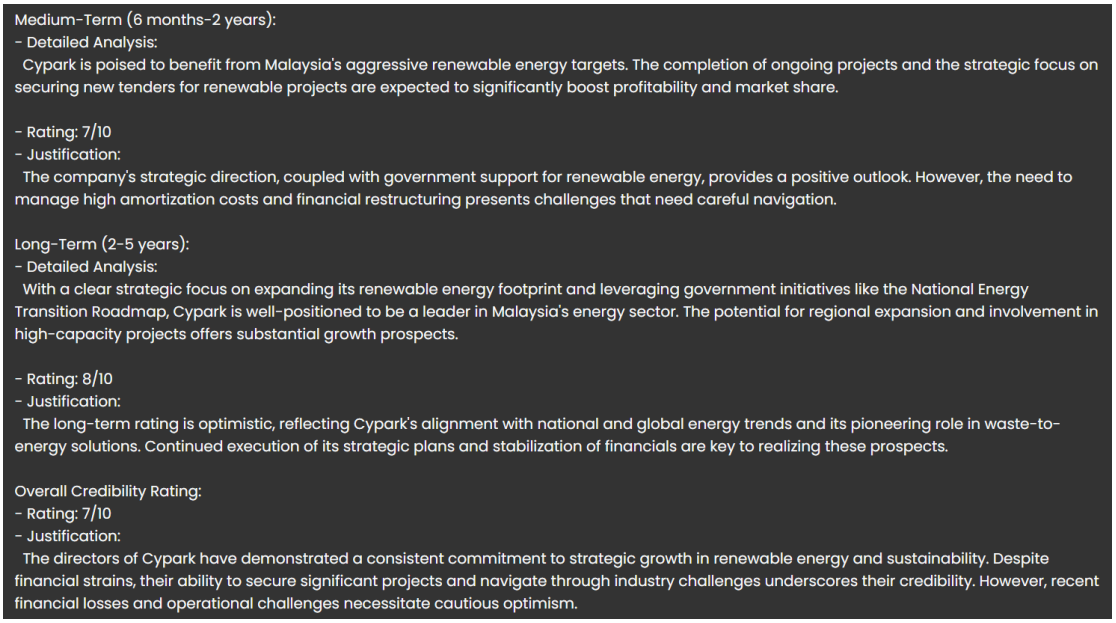


Figure 5.10. Sample Output from System (Cypark with Optimistic) – Part 2

5.5 Implementation Issues and Challenges

Extracting data from annual reports presents significant challenges, primarily due to the varied formats employed by different companies. Creating a standardized method for extracting this information often requires complex, hardcoded solutions that can be difficult to generalize across diverse document types. Additionally, the project's heavy reliance on the GPT model can sometimes result in less-than-optimal responses. The functionality of the system is also vulnerable to disruptions when the GPT server experiences downtime, such as during accidental outages. Consequently, it is crucial to manually verify the GPT-generated data during the preprocessing phase to ensure its accuracy and reliability before proceeding to more detailed analysis. This verification step is vital for maintaining the integrity of the data analysis and ensuring that subsequent insights derived from the system are based on precise and dependable information.

CHAPTER 6

System Evaluation and Discussion

6.1 System Testing and Performance Metrics

As previously mentioned, the reliability of our system - ensuring that the analysis is fact-based rather than speculative - will be rigorously tested. Due to the each analysis required approximately USD 0.10 to 0.12, we will conduct evaluations for 6 companies currently available (50%) in our database which the industry are renewable energy and consumer product. Each company will be analyzed using two distinct analytical perspectives: optimistic and pessimistic. Consequently, this approach results in a total of 12 individual tests, allowing us to thoroughly assess the system's accuracy and consistency across different scenarios and settings.

6.2 Testing Setup and Result

Testing Result Tables for Short-term Prospect.

Green Color meaning that it is actual information otherwise will be Red Color.

Company (Type)	System's Analysis (Short-term)	Actual Information
Solarvest (Optimistic)	- Strong revenue growth and profitability, particularly from the LSS4 projects	- Increase in revenue and profit before tax was mainly contributed by the higher revenue contribution from higher progress of LSS4 projects as compared to the preceding year (from quarterly report)
Solarvest (Pessimistic)	- Financial performance shows significant revenue growth, primarily driven by LSS4 projects - Decrease in revenue in the latest quarter due to	- Increase in revenue and profit before tax was mainly contributed by the higher revenue contribution from higher progress of LSS4 projects as compared to the preceding year (from Quarterly Report)

	the advanced stages of LSS4 projects completion signals	- The Group's revenue for the current financial quarter ended 31 December 2023 decreased by RM27.50 million or 19.66% to RM112.40 million compared to RM139.90 million in the preceding quarter ended 30 September 2023. The marginal decrease in revenue was due to the lower revenue contribution of LSS4 projects. (from Quarterly Report)
Cypark (Optimistic)	- Strategic shift towards Operation & Maintenance (O&M) phases for its renewable energy projects, promising stable, long-term revenue streams	- As we progress, the transition of secured LSS2 projects from the construction phase to the Operation & Maintenance ("O&M") phase is expected to further enhance the Group's revenue streams (From Quarterly Report)
Cypark (Pessimistic)	- financial performance shows a struggle with losses primarily due to high amortization costs and operational challenges in its waste management division - transition of LSS2 projects to the O&M phase	- Accordingly, the Group's loss after tax for 3Q2024 was recorded at RM27.7 million. This was mainly due to the losses recorded by the Waste Management & Waste-To-Energy division in the current quarter due to the adjustment made on the amortization which has been adjusted back to the concession term basis that consistent with the previous year amortization. (From Quarterly Report) - As we progress, the transition of secured LSS2 projects from the

		construction phase to the Operation & Maintenance ("O&M") phase is expected to further enhance the Group's revenue streams (From Quarterly Report)
Samaiden (Optimistic)	- Recent increase in revenue and profit, as reported in the latest quarterly financials	- The Group recorded revenue of RM48.83 million for the current quarter compared to RM40.23 million in the preceding year's corresponding quarter. The increase was due to higher work progress for ongoing projects in the current quarter. (From Quarterly Report)
Samaiden (Pessimistic)	- Positive trend in its recent quarterly performance with increased revenue and profit before taxation	- The Group achieved profit before taxation ("PBT") of RM4.20 million for the current quarter compared to RM3.48 million in the preceding year's corresponding quarter. The higher PBT was in line with the higher revenue recorded and partially offset by increase in professional and stamp duty fees incurred for additional bank facilities. (From Quarterly Report)
Padini (Optimistic)	- Recent quarterly report shows a significant rebound in sales during festive periods, indicating a strong consumer response to strategic marketing and sales initiatives	- In the current quarter under review, revenue and profit before tax increased by approximately RM111.9 million (28.8%) and RM36.0 million (103.1%) respectively, in comparison to the immediately preceding quarter. This was mainly due to heightened sales during festive occasions such as

		Christmas and year end school holidays in the current quarter. (From quarterly report)
Padini (Pessimistic)	- Troubling decline in Padini's financial performance, with a significant drop in profit before tax and a slight decrease in revenue	- Profit before tax reported a quarter-on-quarter decrease of RM26.3million (-27.0%) and a year-on-year decrease of RM55.5 million (-34.4%) respectively. This is partly due to a drop in the gross profit margin from 39% to 37%. (From quarterly report)
Bonia (Optimistic)	- Recent downturn in quarterly performance - Shows resilience with strategic brand collaborations and enhanced digital marketing efforts	- The Group's revenue during the quarter decreased by 3.9% year-on-year to RM117.4 million, primarily due to deceleration in sales within our Singapore market. (From quarterly report) - Looking forward, the company faces economic uncertainties and competitive pressures but remains committed to its growth pillars, including further digital enhancements and international brand collaborations. (From annual report)
Bonia (Pessimistic)	- Recent quarterly report indicates a concerning decline in revenue and profit, particularly in the Singapore market	- The Group's revenue during the quarter decreased by 3.9% year-on-year to RM117.4 million, primarily due to deceleration in sales within our Singapore market. (From quarterly report)
Focus Point (Optimistic)	- Recent quarterly report shows a 5% increase in	- Group revenue at RM260.9 million was 5% higher compared with

	revenue, driven by higher sales in optical products and food & beverage segments	RM248.8 million of the corresponding period ended 31 December 2022. The increase in Group revenue was mainly attributed to higher sales attained by Optical and related products and Food and beverage segments as compared to the corresponding period. (From quarterly report)
Focus Point (Pessimistic)	- Recent quarterly report shows a revenue increase but a decline in profit margins due to higher operating costs from new outlet openings	- Group revenue at RM260.9 million was 5% higher compared with RM248.8 million of the corresponding period ended 31 December 2022. The increase in Group revenue was mainly attributed to higher sales attained by Optical and related products and Food and beverage segments as compared to the corresponding period. The Group recorded a lower profit before tax at RM39.9 million compared with RM47.9 million in the corresponding period mainly due to higher operating costs which included staff costs and rental from the opening of new outlets. (From quarterly report)

Table 6.1. Testing Result for Short-term Prospect.

Testing Result Tables for Medium-term Prospect.

Company (Type)	System's Analysis (Medium-term)	Actual Information
Solarvest (Optimistic)	- Expansion into the CGPP and residential solar solutions	The Group will continue to grow its order book, capitalizing on the imminent awards of up to 800 MW CGPP projects, progressive roll-out of the 10 pilot catalyst projects under the National Energy Transition Roadmap (NETR), among others and remains upbeat on the industry outlook, underpinned by a robust job pipeline in private large-scale power plants, residential, commercial & industrial projects. (From Quarterly Report)
Solarvest (Pessimistic)	- Expanding SOLARVEST's market presence, including international expansions and new sustainable solutions	- The strategic initiatives included leveraging government initiatives, exploring solar leasing programs, and expanding into international markets like the Philippines and Taiwan. (From Credibility Analysis) - The introduction of new sustainable solutions and the enhancement of the Powerbee EV charging network underscore Solarvest's commitment to building a clean energy ecosystem. (From Annual Report)
Cypark (Optimistic)	- Completion of ongoing projects and the strategic focus on securing new	- Opportunities for growth are abundant, with the company focusing on completing ongoing projects and

	tenders for renewable projects	tendering for key RE projects. (From Annual Report)
Cypark (Pessimistic)	- Completion of major projects and potential stabilization of the WTE operations	- Cypark's track record remain as Malaysia's first developer and the only operator of Municipal Solid Waste ("MSW") WTE plant in Ladang Tanah Merah, Negeri Sembilan with capacity of 20MW could stand us in good stead to win more WTE projects in the future, locally and regionally. (From Quarterly Report)
Samaiden (Optimistic)	- Involvement in large-scale projects and its robust order book support sustained growth	- The Corporate Green Power Programme (CGPP) is poised to be a significant driver of job opportunities and a boost to order books in the renewable energy (RE) sector, with an estimated value of contracts ranging between RM2.7 billion to RM3 billion. (From Quarterly Report)
Samaiden (Pessimistic)	- Involvement in government-led renewable energy initiatives and diversification into battery energy storage systems	- The company's business outlook remains optimistic, driven by global demand for sustainable energy and governmental ESG initiatives. Samaiden's experience and strategic ventures in new business verticals, such as Battery Energy Storage Systems and green hydrogen, position it well to capitalize on these trends. (From Annual Report)
Padini (Optimistic)	- Commitment to integrating digital	- Looking forward, Padini is committed to further integrating

	technologies and expanding market presence	digital technologies, expanding its market presence in Southeast Asia, and continuing its focus on sustainability and customer-centric strategies. (From Annual Report)
Padini (Pessimistic)	- Focus on enhancing digital capabilities and expanding market presence	- Looking forward, Padini is committed to further integrating digital technologies, expanding its market presence in Southeast Asia, and continuing its focus on sustainability and customer-centric strategies. (From Annual Report)
Bonia (Optimistic)	- Strategic focus on digital transformation and customer experience enhancement - Continued investment in e-commerce and modernizing physical stores are expected to improve operational efficiencies and customer engagement	- The company's focus on enhancing customer relationships and digital capabilities, alongside expanding e-commerce and modernizing physical stores, has been pivotal. (From Annual Report)
Bonia (Pessimistic)	- Committed to digital growth and international collaborations	- Looking forward, the company faces economic uncertainties and competitive pressures but remains committed to its growth pillars, including further digital enhancements and international brand collaborations. (From Annual Report)
Focus Point (Optimistic)	- Plans for further outlet expansions and	- Future strategies include further expansion of outlet locations,

	enhancements in e-commerce capabilities - Positive economic outlook projected by Bank Negara Malaysia and Focus Point's robust financial position	enhancing e-commerce capabilities, and continuing to invest in technology and workforce development. (From Annual Report) - With gross domestic product (GDP) expected to grow between 4% and 5% in 2024, as projected by Bank Negara Malaysia, the economic momentum is poised to strengthen due to the growing domestic demand. (From Quarterly Report)
Focus Point (Pessimistic)	- Strategic focus on expanding market presence and enhancing e-commerce capabilities - Expected inflationary pressures, might affect consumer spending adversely	- Future strategies include further expansion of outlet locations, enhancing e-commerce capabilities, and continuing to invest in technology and workforce development. (From Annual Report) - With gross domestic product (GDP) expected to grow between 4% and 5% in 2024, as projected by Bank Negara Malaysia, the economic momentum is poised to strengthen due to the growing domestic demand. However, despite this positive outlook, continued inflationary pressures may impact consumer spending. (From Quarterly Report)

Table 6.2. Testing Result for Medium-term Prospect.

Testing Result Tables for Long-term Prospect.

Company (Type)	System's Analysis (Long-term)	Actual Information
Solarvest (Optimistic)	- Strategic direction towards becoming a comprehensive clean energy solutions provider	- Strategic focus on becoming a clean energy specialist, expanding its service offerings, and venturing into new sustainable solutions aligns well with these industry trends (From Credibility Analysis)
Solarvest (Pessimistic)	- Focuses on asset ownership and geographical expansion	- Solarvest's asset ownership and development have also progressed, with LSS4 assets on track and the successful commissioning of new solar plants under the Powervest program. (From Annual Report) - - The strategic initiatives included leveraging government initiatives, exploring solar leasing programs, and expanding into international markets like the Philippines and Taiwan. (From Credibility Analysis)
Cypark (Optimistic)	- Leveraging government initiatives like the National Energy Transition Roadmap	The company plans to leverage its first-mover advantage in floating solar projects and capitalize on the investment opportunities presented by the National Energy Transition Roadmap (NETR).
Cypark (Pessimistic)	- Strong focus on renewable energy and waste-to-energy projects	- Cypark's track record remain as Malaysia's first developer and the only operator of Municipal Solid Waste ("MSW") WTE plant in Ladang Tanah Merah, Negeri

		Sembilan with capacity of 20MW could stand us in good stead to win more WTE projects in the future, locally and regionally. (From Quarterly Report)
Samaiden (Optimistic)	- Commitment to innovation and sustainability, strategic partnerships	- The company's focus on renewable energy solutions, strategic expansion, and operational efficiency, coupled with its commitment to sustainability and innovation, positions it well for continued growth in the dynamic RE sector. (From Annual Report)
Samaiden (Pessimistic)	- Alignment with national sustainability goals and its strategic ventures into new business verticals like green hydrogen	- The company's business outlook remains optimistic, driven by global demand for sustainable energy and governmental ESG initiatives. Samaiden's experience and strategic ventures in new business verticals, such as Battery Energy Storage Systems and green hydrogen, position it well to capitalize on these trends. (From Annual Report)
Padini (Optimistic)	- Governance enhancements and strategic initiatives aimed at leveraging the domestic market and enhancing digital capabilities	- Future growth strategies involve leveraging the domestic market and enhancing digital and operational capabilities to ensure long-term shareholder value. (From Annual Report)
Padini (Pessimistic)	- Focus on sustainability and digital	- Padini's strategic achievements align well with broader retail industry trends, which have increasingly

	transformation, are promising	moved towards digital transformation and globalization. (From Credibility Analysis)
Bonia (Optimistic)	- Promising given Bonia's commitment to innovation and international brand collaborations - Focus on expanding digital capabilities and enhancing the customer experience aligns with future retail trends	- The management's strategic foresight in navigating market dynamics, coupled with a focus on innovation and brand partnerships, positions Bonia for sustained growth despite potential challenges ahead. (From Annual Report) - The company's focus on enhancing customer relationships and digital capabilities, alongside expanding e-commerce and modernizing physical stores, has been pivotal. (From Annual Report)
Bonia (Pessimistic)	- Company's alignment with market trends like digital infrastructure is positive, but operational capacities and technological infrastructures may limit scalability	- The company's focus on enhancing customer relationships and digital capabilities, alongside expanding e-commerce and modernizing physical stores, has been pivotal. (From Annual Report)
Focus Point (Optimistic)	- Commitment to ESG initiatives, strategic location expansions, and ongoing investments in technology and workforce	- The Chairman emphasizes the strategic initiatives for future growth, including expanding market presence, enhancing the brand, and investing in human resources. Focus Point is also committed to ESG initiatives, reflecting a forward-thinking

		approach in governance. (From Annual Report)
Focus Point (Pessimistic)	- Commitment to ESG initiatives and robust governance practices, as outlined in the chairman's statement, aligns with global sustainability trends	- Focus Point is also committed to ESG initiatives, reflecting a forward-thinking approach in governance. The Chairman also outlines the importance of maintaining robust governance practices, with plans for board member rotation and succession planning, ensuring sustainability and ethical business practices. (From Annual Report)

Table 6.3. Testing Result for Long-term Prospect.

All the information generated by the system is based on the inputs provided during the 12 testing scenarios. Having successfully tested 50% of our database with accurate results, we can reasonably assume that the system will maintain this level of reliability across the remaining companies. This consistent performance underlines the system's robustness and effectiveness in delivering dependable analyses.

6.3 Objectives Evaluation

In this project, we have developed an alternative method for making informed stock investments, offering a solution beyond the traditional choices of self-learning or hiring fund managers. Our system significantly reduces the time typically required for self-learning, which involves extensive research into financial markets and investment strategies. Typically, mastering these topics as a beginner is a lengthy and demanding process. Activities such as reading books, subscribing to financial journals, attending seminars, and enrolling in courses can be both time-consuming and costly. However, our system not only analyzes companies but also serves as an educational tool for beginners. It simplifies the learning process by providing direct insights into stock market dynamics and investment strategies, making it easier for newcomers to understand and engage with the market effectively. This dual functionality enhances the accessibility of stock market investment for beginners, making it a practical and

valuable resource for their financial education and investment activities. Another option for making informed stock investments is hiring a fund manager. Fund managers are responsible for implementing the fund's strategy and managing its trading activities. However, hiring a fund manager can be expensive, with costs varying based on their experience and qualifications. It's important to note that none of the methods guarantees performance or results, and there's a risk of financial loss regardless of the amount paid to a fund manager, given that returns are not guaranteed. In contrast, our system focuses on reducing monetary costs. Instead of incurring the substantial expense of hiring a fund manager, we offer a pay-as-you-go model. This approach allows users to analyse individual companies at a fraction of the cost, paying only the fees required for GPT API usage.

Additionally, value investment is typically a time-consuming task for stock market professionals and experts. However, features like "Relevant Data," "Malaysia Trends," and our analyzer can significantly speed up this process. The "Relevant Data" function consolidates all necessary data, performs preprocessing to organize it tidily, and compiles everything onto a single webpage. This integration eliminates the need for users to toggle between multiple websites or PDFs to gather different pieces of information. Similarly, "Malaysia Trends" and the analyzer streamline the process by providing quick access to relevant market trends and analytical insights, all in one place, enhancing efficiency for users.

Moreover, our system provides recommendations over a range of timeframes with supporting evidence, such as short-term (3 months to 6 months), medium-term (6 months to 2 years), and long-term (2 years to 5 years), rather than merely predicting tomorrow's stock price. This approach aligns more closely with the principles of fundamental analysis, as we focus on assessing the company's foundational aspects such as financial health, business model, industry position, and market conditions. This deeper analysis helps in understanding the long-term potential and sustainability of a company, which is crucial for making informed investment decisions.

CHAPTER 7

Conclusion and Recommendation

7.1 Conclusion

Investing serves as a beneficial means to supplement income. However, delving into stock market investments often requires significant time and effort in researching companies. Hence, an easier way is to use the prediction tools, but most of the tools are quantitative based, which can raise the issue that the stock market is not merely infected by the numbers but by unpredictable variables like economic and political changes. Therefore, a qualitative based prediction system can resolve the issue raised by quantitative data. This project introduces such a system, emphasizing qualitative data to address the limitations of quantitative-only approaches.

This project utilizes financial information such as annual reports of the company, related news, and government policies to analyze the performance of the company in the future. The special feature of the system is to create a user-friendly environment for people to interact with the GPT model easily and retrieve optimized results at the same time. Moreover, this system enhances the GPT's capabilities in financial analysis, overcoming ChatGPT's limitation of not having real-time financial data access.

7.2 Recommendation

Due to the time limitations, there is quite a lot of interesting topics to this project.

i. Revenue in Depth

This concept aims to enhance our analysis of companies by focusing on internal factors alongside external ones such as competitors and government policies. Specifically, by analyzing a company's revenue streams, we can identify key customers and understand the nature of their relationship with the company. This approach allows us to gain valuable investment insights by exploring how these relationships impact the company's revenue. If a significant relationship exists, it suggests that the customer base directly affects the financial health of the company,

providing crucial information for investors about the company's stability and growth prospects.

For example, consider Company A, a chip manufacturer, and Company B, a phone maker. Analysis of Company A's revenue reveals that 91% of its income is derived from sales to Company B. Consequently, Company A's financial performance is closely tied to Company B's success. If Company B's phones sell well, leading to increased profits, it is likely to place larger orders with Company A to meet the demand. This in turn would boost Company A's revenue.

ii. Comparable

This function enables the comparison of multiple companies across different industries. Currently, our system focuses on comparing competitors within the same industry, but this new feature expands our analytical capabilities to encompass cross-industry comparisons.

iii. AI-Youtuber

Since GPT's ability to interpret or analyze data from multiple perspectives has been demonstrated, we can envision creating an AI version of a YouTuber. By feeding GPT the transcript from a YouTuber’s videos, it can learn how the YouTuber analyzes data and presents information. Subsequently, GPT can emulate the YouTuber’s style and approach to provide users with analyses of companies.

REFERENCES

[1] D. P. Gandhmal and K. Kumar, "Systematic analysis and review of Stock Market Prediction Techniques," *Computer Science Review,* vol. 34, p. 100190, 2019. doi:10.1016/j.cosrev.2019.08.001

[2] A. Thakkar and K. Chaudhari, "Fusion in stock market prediction: A decade survey on the necessity, recent developments, and potential future directions," *Information Fusion,* vol. 65, pp. 95–107, 2021. doi:10.1016/j.inffus.2020.08.019

[3] W. Jiang, "Applications of deep learning in stock market prediction: Recent progress," *Expert Systems with Applications,* vol. 184, p. 115537, 2021. doi:10.1016/j.eswa.2021.115537

[4] S. N. Balaji, P. V. Paul, and R. Saravanan, "Survey on sentiment analysis based stock prediction using Big Data Analytics," *2017 Innovations in Power and Advanced Computing Technologies (i-PACT),* 2017. doi:10.1109/ipact.2017.8244943

[5] R. Andrawos, "NLP in Stock Market Prediction: A Review - Researchgate," NLP in Stock Market Prediction: A Review, https://www.researchgate.net/publication/361164679_NLP_in_Stock_Market_Prediction_A_Review (accessed Sep. 11, 2023).

[6] A. Lopez-Lira and Y. Tang, "Can CHATGPT forecast stock price movements? return predictability and large language models," SSRN, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4412788 (accessed Sep. 9, 2023).

[7] R. Steinert and S. Altmann, "Linking microblogging sentiments to stock price movement: An application of GPT-4," arXiv.org, https://arxiv.org/abs/2308.16771 (accessed Sep. 10, 2023).

[8] "Who we are," Who We Are, https://www.bursamalaysia.com/about_bursa/about_us/who_we_are (accessed Dec. 7, 2023).

[9] G. Lawton, "What is Generative Ai? everything you need to know," Enterprise AI, https://www.techtarget.com/searchenterpriseai/definition/generative-AI (accessed Dec. 7, 2023).

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Trimester 3, Year 3	**Study week no.: 2**
Student Name & ID: Teo Teck Wan (20ACB01750)	
Supervisor: Prof Ts Dr. Liew Soung Yue	
Project Title: AI–Assisted Value Investment in Malaysian Stock Market with Generative AI	

1. WORK DONE

No.

2. WORK TO BE DONE

Exploring the government policies that required into the system.

3. PROBLEMS ENCOUNTERED

No.

4. SELF EVALUATION OF THE PROGRESS

Good.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Trimester 3, Year 3	**Study week no.: 4**
Student Name & ID: Teo Teck Wan (20ACB01750)	
Supervisor: Prof Ts Dr. Liew Soung Yue	
Project Title: AI–Assisted Value Investment in Malaysian Stock Market with Generative AI	

1. WORK DONE

Choose the department of Transport, Education, Science, Technology, and Innovation, and Economy.
Choose the Budget 2024, Madani Economy, New Industrial Master Plan 2030, RMK12, New Energy Transition Roadmap.
Start to scrap the information from these official website.

2. WORK TO BE DONE

Scrap all the policies mentioned above into our database.

3. PROBLEMS ENCOUNTERED

No.

4. SELF EVALUATION OF THE PROGRESS

Good.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Trimester 3, Year 3	**Study week no.: 6**
Student Name & ID: Teo Teck Wan (20ACB01750)	
Supervisor: Prof Ts Dr. Liew Soung Yue	
Project Title: AI–Assisted Value Investment in Malaysian Stock Market with Generative AI	

1. WORK DONE

Add “Progress News” to improve the prediction of Malaysia Trends in 2024.
Start to perform the prediction from all the policies and progress news.

2. WORK TO BE DONE

Prediction is done by using the ChatGPT instead of the GPT API because of the memory features that ChatGPT have.

3. PROBLEMS ENCOUNTERED

No

4. SELF EVALUATION OF THE PROGRESS

Good.

Supervisor’s signature

Student’s signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Trimester 3, Year 3	**Study week no.: 8**
Student Name & ID: Teo Teck Wan (20ACB01750)	
Supervisor: Prof Ts Dr. Liew Soung Yue	
Project Title: AI–Assisted Value Investment in Malaysian Stock Market with Generative AI	

1. WORK DONE

Start to perform the credibility analysis and competitor analysis.

2. WORK TO BE DONE

Two analysis is done for all the companies.

3. PROBLEMS ENCOUNTERED

No.

4. SELF EVALUATION OF THE PROGRESS

Good.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Trimester 3, Year 3	**Study week no.: 10**
Student Name & ID: Teo Teck Wan (20ACB01750)	
Supervisor: Prof Ts Dr. Liew Soung Yue	
Project Title: AI–Assisted Value Investment in Malaysian Stock Market with Generative AI	

1. WORK DONE

Perform the optimistic version of analyzer and pessimistic version as well.

2. WORK TO BE DONE

Coding and set the GPT to view data optimistically or pessimistically.

3. PROBLEMS ENCOUNTERED

No.

4. SELF EVALUATION OF THE PROGRESS

Good.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Trimester 3, Year 3	**Study week no.: 12**
Student Name & ID: Teo Teck Wan (20ACB01750)	
Supervisor: Prof Ts Dr. Liew Soung Yue	
Project Title: AI–Assisted Value Investment in Malaysian Stock Market with Generative AI	

1. WORK DONE

Add on the quarterly report.
Increase the company size in database from 3 to 11.

2. WORK TO BE DONE

Adding the quarterly report into the analysis.
Database available company increased to 11.

3. PROBLEMS ENCOUNTERED

No.

4. SELF EVALUATION OF THE PROGRESS

Good.

Supervisor's signature

Student's signature

POSTER


AI-ASSISTED
VALUE INVESTMENT IN
MALAYSIAN STOCK MARKET
WITH GENERATIVE AI
New Qualitative-Based Recommendation System to Stock Market
We achieve to
Offer an alternative method for making informed investment
Accelerate the process of value investment
Provide from 3 months to 5 years recommendation ratings with evidences
Our system:
• Available in Malaysian Stock Market
• Information Retrieval Tools
• Utilizes GPT-4.0 Capabilities
• Analyze Different Point of Views
• Select Company & Get All Analysis
We solved
Time Consuming in Value Investment
We assisted
Beginner in Stock Market to Invest Properly
BACHELOR OF COMPUTER SCIENCE (HONS)
By Teo Teck Wan
UTAR
UNIVERSITI TUNKU ABDUL RAHMAN

PLAGIARISM CHECK RESULT

Document Viewer

Turnitin Originality Report

Processed on: 26-Apr-2024 10:13 +08
ID: 2362125590
Word Count: 14767
Submitted: 1

AI–Assisted Value Investment in Malaysian Sto... By Teo Teck Wan

Similarity Index	Similarity by Source	
13%	Internet Sources:	10%
	Publications:	3%
	Student Papers:	6%

exclude quoted | include bibliography | excluding matches < 8 words | mode: quickview (classic) report | print | download

3% match (student papers from 24-Apr-2023)
Submitted to Universiti Tunku Abdul Rahman on 2023-04-24

2% match (Dattatray P. Gandhmal, K. Kumar. "Systematic analysis and review of stock market prediction techniques", Computer Science Review, 2019)
Dattatray P. Gandhmal, K. Kumar. "Systematic analysis and review of stock market prediction techniques", Computer Science Review, 2019

1% match (student papers from 17-Apr-2017)
Submitted to Universiti Tunku Abdul Rahman on 2017-04-17

1% match (Internet from 20-Jan-2024)
https://focuspoint.listedcompany.com/newsroom/FPHB_-_Quarterly_Financial_Result_-_30.09.2023_20231123.pdf

<1% match (student papers from 08-Sep-2022)
Submitted to Universiti Tunku Abdul Rahman on 2022-09-08

<1% match (student papers from 25-Apr-2023)
Submitted to Universiti Tunku Abdul Rahman on 2023-04-25

<1% match (student papers from 22-Apr-2022)
Submitted to Universiti Tunku Abdul Rahman on 2022-04-22

<1% match (student papers from 21-Apr-2022)
Submitted to Universiti Tunku Abdul Rahman on 2022-04-21

<1% match (student papers from 08-Sep-2023)
Submitted to Universiti Tunku Abdul Rahman on 2023-09-08

APPENDIX

Universiti Tunku Abdul Rahman			
Form Title : Supervisor's Comments on Originality Report Generated by Turnitin for Submission of Final Year Project Report (for Undergraduate Programmes)			
Form Number: FM-IAD-005	Rev No.: 0	Effective Date: 01/10/2013	Page No.: 1of 1

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

Full Name(s) of Candidate(s)	Teo Teck Wan
ID Number(s)	20ACB01750
Programme / Course	BACHELOR OF COMPUTER SCIENCE (HONOURS)
Title of Final Year Project	AI-Assisted Value Investment in Malaysian Stock Market with Generative AI

Similarity	**Supervisor's Comments (Compulsory if parameters of originality exceeds the limits approved by UTAR)**
Overall similarity index: 13 % **Similarity by source** Internet Sources: 10 % Publications: 3 % Student Papers: 6 %	Within the required range.
Number of individual sources listed of more than 3% similarity: 0	Within the required range.
Parameters of originality required and limits approved by UTAR are as Follows: **(i) Overall similarity index is 20% and below, and** **(ii) Matching of individual sources listed must be less than 3% each, and** **(iii) Matching texts in continuous block must not exceed 8 words** ***Note: Parameters (i) – (ii) shall exclude quotes, bibliography and text matches which are less than 8 words.***	

Note Supervisor/Candidate(s) is/are required to provide softcopy of full set of the originality report to Faculty/Institute

Based on the above results, I hereby declare that I am satisfied with the originality of the Final Year Project Report submitted by my student(s) as named above.

Signature of Supervisor

Name: Liew Soung Yue

Date: 26/4/2024

Signature of Co-Supervisor

Name: _____

Date: _____

UNIVERSITI TUNKU ABDUL RAHMAN

FACULTY OF INFORMATION & COMMUNICATION TECHNOLOGY
(KAMPAR CAMPUS)

CHECKLIST FOR FYP2 THESIS SUBMISSION

Student Id	20ACB01750
Student Name	Teo Teck Wan
Supervisor Name	Prof Ts Dr. Liew Soung Yue

TICK (√)	DOCUMENT ITEMS Your report must include all the items below. Put a tick on the left column after you have checked your report with respect to the corresponding item.
√	Title Page
√	Signed Report Status Declaration Form
√	Signed FYP Thesis Submission Form
√	Signed form of the Declaration of Originality
√	Acknowledgement
√	Abstract
√	Table of Contents
√	List of Figures (if applicable)
√	List of Tables (if applicable)
√	List of Symbols (if applicable)
√	List of Abbreviations (if applicable)
√	Chapters / Content
√	Bibliography (or References)
√	All references in bibliography are cited in the thesis, especially in the chapter of literature review
√	Appendices (if applicable)
√	Weekly Log
√	Poster
√	Signed Turnitin Report (Plagiarism Check Result - Form Number: FM-IAD-005)
√	I agree 5 marks will be deducted due to incorrect format, declare wrongly the ticked of these items, and/or any dispute happening for these items in this report.

*Include this form (checklist) in the thesis (Bind together as the last page)

I, the author, have checked and confirmed all the items listed in the table are included in my report.

_____
(Signature of Student)
Date: 26/4/2024

APPENDIX