Accelerated Personalized Stock Sentiment Analysis: Leveraging LLMs for YouTuber Content and News Articles

BY

Sim Kah Hoe

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)

June 2024

UNIVERSITI TUNKU ABDUL RAHMAN

REPORT STATUS DECLARATION FORM

Title: Accelerated Personalized Stock Sentiment Analysis: Leveraging LLMs for YouTuber Content and News Articles

Academic Session: June 2024

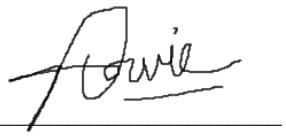
I Sim Kah Hoe

(CAPITAL LETTER)

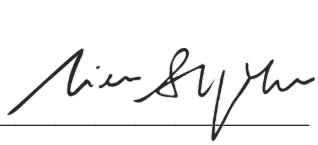
declare that I allow this Final Year Project Report to be kept in Universiti Tunku Abdul Rahman Library subject to the regulations as follows:

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:
216, Jalan Seri Binjai 6, Taman Seri Binjai, 70100, Seremban, Negeri Sembilan

Liew Soung Yue

Supervisor's name

Date: **12 September 2024**

Date: 13/9/2024

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

UNIVERSITI TUNKU ABDUL RAHMAN

Date: 12 September 2024

SUBMISSION OF FINAL YEAR PROJECT /DISSERTATION/THESIS

It is hereby certified that Sim Kah Hoe (ID No: 21ACB06036) has completed this final year project entitled "Accelerated Personalized Stock Sentiment Analysis: Leveraging LLMs for YouTuber Content and News Articles" 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 project / dissertation/ thesis* in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,

_____

(*Sim Kah Hoe*)

DECLARATION OF ORIGINALITY

I declare that this report entitled "**ACCELERATED PERSONALIZED STOCK SENTIMENT ANALYSIS: LEVERAGING LLMS FOR YOUTUBER CONTENT AND NEWS ARTICLES**" 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 : Sim Kah Hoe

Date : 12 September 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 an innovative AI project. It is my first step to establish a career in large language model field. A million thanks to you.

I must say thanks to my parents and my family for their love, unyielding support, and continuous encouragement throughout the course.

ABSTRACT

This project examines the transformative potential of natural language processing (NLP), specifically through the use of ChatGPT, in the realm of stock investment. The primary goal is to create a dynamic, user-focused AI-driven system that provides investors with real-time insights, tailored analyses, and enhanced decision support for the stock market. The project encompasses a broad scope, including data integration, model adaptation, system development, performance evaluation, and ongoing improvements. Central to this effort is the use of ChatGPT 4.0. This interdisciplinary approach highlights the project's dedication to bridging the gap between AI and stock investment. The project's innovation lies in its ability to enhance decision-making support for investors by leveraging AI's NLP capabilities to facilitate intuitive interactions and deliver real-time insights. The iterative learning process ensures that the system remains adaptable and continuously improves, while comprehensive documentation aids in knowledge sharing within the financial sector. In essence, this research represents a significant advancement toward democratizing stock investment, making it more accessible and data-driven. By leveraging ChatGPT and cutting-edge technologies, the project provides investors with a valuable tool for navigating the complexities of the stock

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

NLP

Natural Language Processing

LLM

Large Language Model

AI

Artificial Intelligence

LSTM

Long-Short-Term Memory

API

Application Programming Interface

MDA

Management's Discussion and Analysis

LIME

Local Interpretable Model-agnostic Explanations

SHAP

SHapley Additive exPlanations

Chapter 1

Introduction

Investor sentiment significantly impacts stock market fluctuations. News reports, social media discussion, and the general market mood can heavily influence stock prices. Accurately accessing this sentiment is crucial for making informed investment decisions. Large Language Models (LLMs), such as OpenAI's ChatGPT, provide a promising solution in this area. These AI models excel at processing extensive amounts of unstructured textual data, including financial news and social media conversations. By analyzing this data, LLMs can extract sentiment and offer valuable insights for investors. Integrating sentiment analysis with stock investing introduces a novel and potentially transformative approach. With ChatGPT, investors can analyze diverse data sources to identify trends and make data-driven decisions. Real-time sentiment analysis, personalized investment guidance could become powerful tools for investors. As the financial landscape evolves, LLMs like ChatGPT have the potential to revolutionize stock investment, making data-driven insights a cornerstone of investment strategies and helping investors navigate market complexities to achieve superior outcomes.

1.1 Problem Statement and Motivation

Current sentiment analysis methods for the stock market have notable limitations. Traditional approaches often struggle with the nuances and context of financial language, leading to potentially inaccurate or incomplete assessments. This difficulty in interpreting textual data from news and social media can result in flawed sentiment information, which may negatively impact investor decision making and reduce returns. The dynamic nature of the stock market demands timely and informed decisions, yet the lack of real-time sentiment analysis tools hampers the ability to respond swiftly to changing conditions. This can result in missed opportunities or misinterpretations of market trends.

As the financial landscape becomes increasingly complex, the volume of unstructured text data grows, presenting challenges in extracting actionable insights. Traditional methods often fall short, leading to suboptimal investment strategies. This project aims to address these issues by leveraging ChatGPT's advanced capabilities. The goal is to provide investors with a sophisticated, real-time sentiment analysis tool that offers personalized investment guidance. By bridging the gap between data influx and effective decision-making, this solution will equip investors with deeper insights and more agile strategies in an era driven by data.

1.2 Objectives

The ever-increasing volume of financial information presents a significant challenge for investors. Navigating through countless news articles and extensive YouTube analyses can be both time-consuming and laborious. This project aims to address these challenges by leveraging Large Language Models (LLMs) and sentiment analysis to streamline the investment research process and enhance informed decision-making. The primary objective of this project is to tackle the central pain point of information overload for investors. Currently, investors spend considerable time manually searching for relevant news and analyzing extensive YouTube content. This project seeks to alleviate this burden by automatically parsing and analyzing vast amounts of data, including news articles and YouTube transcripts, thus providing access to the latest information and expert insights more efficiently. In addition to managing information overload, understanding the overall sentiment towards a particular stock can be complex. This project focuses on utilizing sentiment analysis techniques to deliver a quick and accurate assessment of prevailing market sentiment, enabling investors to make well-informed decisions in a timely manner.

To offer a more comprehensive view, the project includes several key features:

1. **Identifying Potential Sentiment Shifts**: By analyzing the data, the project will highlight potential changes in sentiment over time, helping investors anticipate market movements and adopt a proactive investment strategy.
2. **Incorporating Historical Context**: The project will integrate historical stock performance data to provide a more thorough understanding of a stock's trajectory. This historical context will support more accurate sentiment interpretation and forecasting.

3. **Generating Data-Driven Forecasts**: Leveraging LLMs, the project will produce predictive insights for future stock prices, equipping investors with valuable information to guide their investment decisions.
4. **Personalized Investment Recommendations**: The project will utilize LLM capabilities to tailor investment advice based on individual risk tolerance and goals, ensuring alignment with each investor's unique financial situation.
5. **Creating a Sentiment Score**: An LLM-powered sentiment score will be developed to provide a clear, quantifiable indicator of investor sentiment, simplifying the decision-making process.

By achieving these objectives, this project aims to offer investors a powerful suite of tools that streamline their research, provide valuable insights, and empower them to make informed, data-driven investment decisions. Ultimately, this approach is designed to enhance the efficiency and potential profitability of the investment experience.

1.3 Project Scope and Direction

This project aims to investigate the potential of sentiment analysis for forecasting stock prices using qualitative data. The focus will be on analyzing user-selected YouTube content related to stock analysis, alongside current company news articles and historical stock data. Users will have the ability to choose their preferred stock analysis YouTubers and add them to a watchlist. Sentiment analysis will be performed on the content from these selected YouTubers, as well as on news articles related to companies from the past two weeks. The objective is to generate comparative reports that explore the sentiment derived from both sources and assess its correlation with stock price movements and historical trends. The effectiveness of sentiment-based predictions will be evaluated by comparing them against historical stock data, providing insights into the reliability and potential of qualitative data for stock price forecasting. This comprehensive analysis will highlight the strengths and limitations of sentiment analysis in predicting stock prices and offer recommendations for future exploration in this field.

1.4 Contributions

This project aims to transform how investors conduct research and make decisions by harnessing the capabilities of Large Language Models (LLMs) and sentiment analysis. It addresses the challenge of information overload in today's data-intensive environment by automating the parsing and analysis of vast amounts of data from selected YouTube stock analyses and current news articles. This automated approach simplifies access to up-to-date information and insights, significantly reducing the time and effort required for manual data review. Sentiment analysis will provide a quick and precise evaluation of market sentiment related to specific stocks, enabling investors to make well-informed and timely decisions. The project goes beyond basic sentiment analysis by offering actionable insights and detecting potential sentiment shifts that could signal market changes or emerging trends. By incorporating historical stock data, the project will offer a deeper understanding of a stock's performance and its potential future trajectory. Additionally, the use of LLMs will facilitate the generation of data-driven forecasts and personalized investment recommendations, tailored to individual risk tolerance and investment goals. This comprehensive approach is designed to enhance the efficiency and effectiveness of investors' research processes, ultimately improving their decision-making capabilities and potential profitability.

1.5 Report Organization

This report is structured into five chapters. Chapter 1 introduces the project by detailing the problem statement, motivation, project scope, objectives, contributions, and providing an overview of the industry background. Chapter 2 presents a literature review that examines investment theories and recent advancements in machine learning. Chapter 3 discusses the overall methodology and design of the system used in this project. Chapter 4 describes the system design and outlines the processes involved. Chapter 5 focuses on the implementation details of the system. Chapter 6 evaluates the system's reliability through various assessment methods. Finally, Chapter 7 summarizes the project and outlines potential areas for future work.

Chapter 2

Literature Review

2.1 Previous Works on Application of Sentiment Analysis for Annual Report and News

2.1.1 Using Sentimental Analysis in Prediction of Stock Market Investment

From [1] S. K. Khatri and A. Srivastava. Sentiment analysis, a prominent tool across various industries, including the stock market, plays a pivotal role in extracting sentiments from user comments to gauge public mood, which can have a significant impact on stock prices. Artificial neural networks (ANN) have emerged as valuable assets, being trained with sentiment analysis results and stock prices to predict future investment returns. Researchers have adeptly harnessed data from social media platforms such as Twitter and StockTwits, applying sentiment analysis to create overall sentiment indices as inputs for ANN prediction models. This approach, combining sentiment scores with market data and feeding them into ANN systems, enhances the accuracy of investment predictions. Sentiment analysis has also been instrumental in generating predictors for investment, specifically focusing on sentiment indicators sourced from Twitter and the presence of monetary terms. Previous studies have consistently emphasized the utility of social data, such as tweets and posts from platforms like Twitter and Facebook, in predicting consumer sentiments towards companies and forecasting product sales. Furthermore, sentiment analysis conducted on stock-linked tweets has revealed noteworthy associations between sentiment indicators and unusual returns and trading volume. The microblogging service StockTwits has served as a valuable resource for sentiment analysis, employing machine learning algorithms to classify post sentiments and create predictive models for market directions. Moreover, sentiment analysis has demonstrated its prowess in accurately forecasting the closing index of specific companies, exemplified by the high accuracy achieved for Tata Services. Additionally, researchers have employed sentiment analysis to construct a social behavior graph, uncovering correlations between trading activities and stock price fluctuations. Lastly, sentiment analysis utilizing

SentiWordNet has elucidated the intricate relationship between market trends and product popularity and ratings. These findings collectively underscore the substantial impact of sentiment analysis on enhancing decision-making processes in the stock market. The following figure 2.1 are the results obtained.

$$f(x) = \frac{1}{1+e^{-x}} \quad (3)$$

The neural network used in our research work is feed forward network which enables a neuron to send its output to another neuron as input and output of these neurons are transferred to hidden layer as input and whose output finally become the input of output layer.

Table II shows the output value calculated for each company with the mean square error.

TABLE 2: RESULTS OBTAINED

Company Name	No. of Neurons	MSE
Apple	10	0.14
Google	12	0.27
Microsoft	10	0.18
Oracle	9	0.22
Facebook	15	0.28

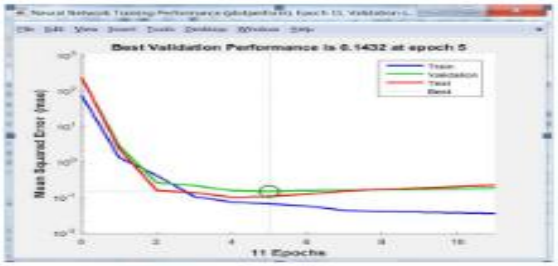


Fig. 1. Implementation of Apple's data

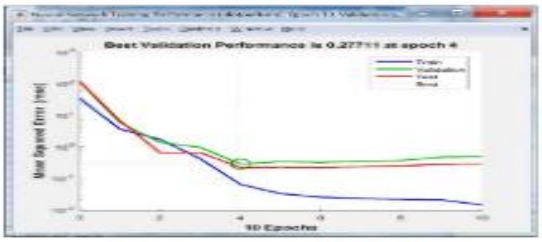


Fig. 2. Implementation of Google's data

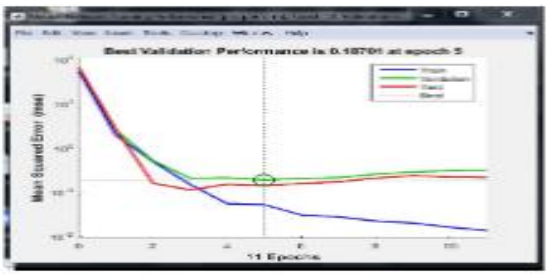


Fig. 3. Implementation of Microsoft's data

Fig. 4. Implementation of Oracle's data

Fig. 5. Implementation of Facebook's data

Figure 2.1: Implementation of Different Sources

2.1.2 Enhancing Stock Investment Strategies through Annual Report Analysis with Large Language Models

[2] U. Gupta. The analysis of annual reports from publicly listed companies has long been a cornerstone of financial assessment, offering valuable insights into their financial health and potential impact on stock prices. Financial experts have traditionally excelled in extracting critical information from these reports, honing their skills through years of practice and experience. However, the advent of Large Language Models (LLMs) such as GPT-3.5 has ushered in a new era. These advanced models possess unparalleled Natural Language Understanding capabilities, enabling them to comprehensively analyze annual reports and generate valuable insights. Researchers have explored the application of LLMs in the financial domain, particularly in predicting stock performance. Machine Learning models trained with LLM-generated outputs as features have demonstrated the potential to predict the best-performing stocks over the upcoming year, offering a data-driven approach to stock selection. Moreover, studies have delved into the impact of including transaction costs in stock selection strategies based on news headlines, highlighting the notable influence these costs can have on net returns. Additionally, a shift towards longer prediction horizons, such as six or twelve months, has been observed, with research indicating that these extended timeframes can yield substantial positive net gains, providing investors with more comprehensive and beneficial insights for their decision-making processes. Figure 2.1 shows neural models used for financial analysis. The first is a character-based language model trained on financial news articles, utilizing a character embedding followed by an Long Short-Term Memory (LSTM) layer for text generation. The second model, designed for stock price prediction, also employs an LSTM but with an additional fully connected layer. It uses the text generated by the language model as input and predicts the probability of a positive stock price change. Both models are trained using different optimization techniques over several epochs to capture financial patterns and make predictions.

2.1.3 Enhancing Sentiment Analysis based Investment by Large Language Models in Japanese Stock Market

From [3] M. Nakano and T. Yamaoka. The paper introduces and evaluates two distinct sentiment analysis approaches for their effectiveness in predicting stock price movements. The first approach employs a traditional word frequency method, utilizing a Bag-of-Words approach to count the number of positive and negative words within news headlines. This approach aims to quantify sentiment based on the prevalence of positive and negative language in news articles. In contrast, the second approach leverages the advanced capabilities of ChatGPT, a state-of-the-art language model. Instead of relying on predefined word lists, this approach directly interacts with ChatGPT to inquire about the positive or negative impact of news articles on stock price movements. By engaging ChatGPT in conversation, it seeks nuanced insights beyond mere word counting. To assess the effectiveness of these sentiment analysis approaches, the paper utilizes sentiment scores generated through both methods. These sentiment scores serve as potential indicators of market sentiment. To validate their utility, the paper employs rigorous statistical tests and conducts investment simulations. The results of these analyses reveal that the sentiment analysis strategy utilizing ChatGPT consistently outperforms conventional benchmarks, such as the market index, as well as traditional investment strategies like mean-variance and risk parity portfolios. This finding underscores the potential of advanced natural language processing models like ChatGPT in enhancing predictive analytics in the realm of stock market investments, offering a more sophisticated and effective approach to sentiment analysis with tangible benefits for investors and portfolio managers. The figures below show a list positive and negative words selected by ChatGPT, which can help with sentiment analysis.

Figure 2.2: D_p positive words selected by ChatGPT

下落 (decline)	続落 (continued decline)	売り (sell)	売却 (sale)
値下がり (price decline)	下値 (lower price)	急落 (plunge)	減益 (decrease in profit)
下方 (downward)	逆風 (headwind)	マイナス (negative)	空売り (short sale)
売り出し (offering)	売ら (sold)	下回る (below)	ネガティブ (negative)

Figure 2.3: D_n negative words selected by ChatGPT

2.1.4 Which Artificial Intelligence Algorithm Better Predicts the Chinese Stock Market?

[4] "Which Artificial Intelligence Algorithm Better Predicts the Chinese Stock Market?" July 2018. This study examines the effectiveness of various artificial intelligence (AI) algorithms in predicting stock market behavior, with a particular focus on the Chinese stock market. The introduction highlights the limitations of traditional stock market prediction methods, such as autoregressive integrated moving average (ARIMA) models, support vector machines, and different neural network architectures, due to the nonlinear and complex nature of market behavior. Recent advancements in artificial intelligence, especially deep learning (DL), offer new possibilities for enhancing prediction accuracy. Deep learning, utilizing multi-layer neural networks, can uncover complex patterns from large datasets without needing prior predictive information. Central to this research is a deep-learning-based stock index futures prediction model that integrates an autoencoder and a restricted Boltzmann machine. The model uses high-frequency transaction data from the CSI 300 futures contract (IF1704) collected from February 20 to April 20, 2017, including one-minute interval data on opening prices, highest prices, lowest prices, closing prices, trade volumes, and opening interests. The study compares the performance of this deep learning model with three traditional artificial neural networks: the back propagation neural network (BP), the extreme learning machine (ELM), and the radial basis function neural network (RBF), across small, medium, and large datasets to evaluate data size impact on predictive performance. The results indicate that the deep learning model significantly surpasses traditional methods in all metrics, including root-mean-square error (RMSE), mean absolute percentage error (MAPE), and directional predictive accuracy. The deep learning model shows a lower RMSE and higher directional accuracy, demonstrating improved predictive performance as dataset size increases. The study concludes that deep learning models offer substantial advantages in stock market prediction, despite their higher computational demands, and advocates for further research incorporating complex network theories and advanced methodologies to enhance predictive capabilities.

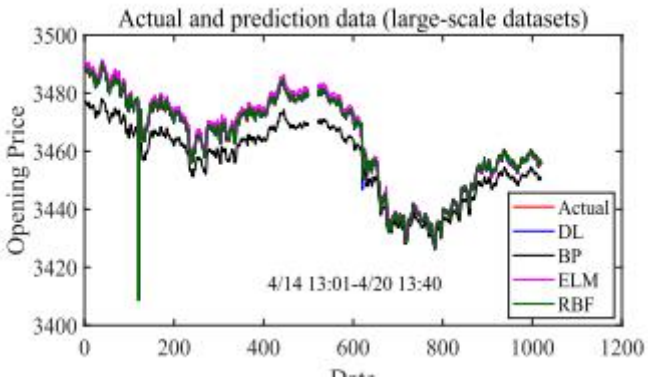


Figure 2.4 Actual and prediction data of the CSI 300 futures contact (Large-scale datasaets)

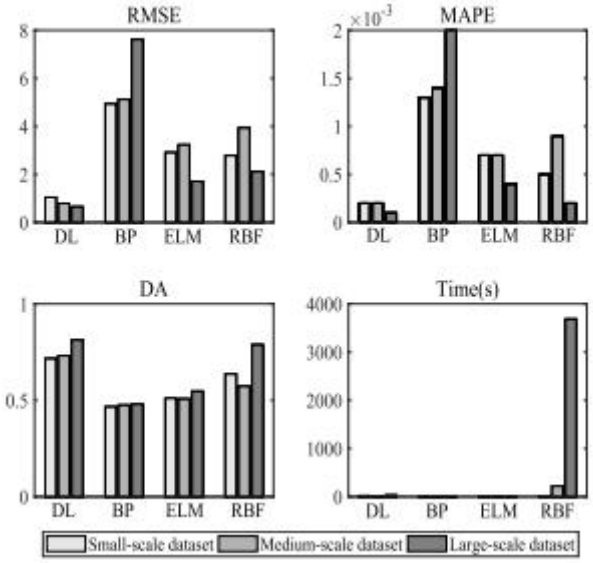


Figure 2.5 Predictive performance comparison of different scale datasets

2.1.5 Using AI to Make Predictions on Stock Market

This paper investigates the use of machine learning for predicting short-term stock prices. It highlights the crucial role of stock trading in finance and the increasing relevance of machine learning in this domain, noting that some studies have shown promising results in stock price forecasting.

The authors provide an overview of previous research on stock price prediction using machine learning, discussing various approaches, time frames, and predictors. While advancements have been made with accuracy rates exceeding 70% in some studies, the authors point out that these models often depend on company-specific information, which may not be suitable for short-term predictions.

Focusing on short-term predictions with time series data from 82 randomly selected NYSE stocks, the study utilizes the Alpha Vantage API for daily stock data, including open, high, low, close prices, and volume. The authors preprocess this data using complex models before applying machine learning algorithms, achieving nearly 70% accuracy in predicting price trends. They note that adding factors like market news could further improve the model's performance.

The paper concludes by emphasizing the significance of the results, despite the modest accuracy. The authors suggest exploring additional factors and neural network-based models for future research to enhance prediction accuracy and effectiveness.

(m, n)	(20, 5)	(20, 3)	(10, 3)	(10, 1)	(5, 1)
Logistic Regression	51.26%	51.04%	52.37%	47.97%	48.08%
Bayesian Network	50.84%	50.97%	48.52%	47.09%	46.87%
Simple Neural Network	47.06%	45.93%	44.66%	42.14%	42.83%
SVM with rbf kernel	46.22%	44.79%	43.38%	41.33%	41.01%

Figure 2.6 Test error rates for different models

2.1.6 Application of Artificial Intelligence in Stock Market Forecasting: A Critique, Review, and Research Agenda

[6] The study titled "Application of Artificial Intelligence in Stock Market Forecasting: A Critique, Review, and Research Agenda" explores the use of artificial intelligence (AI) in predicting stock market trends. The unpredictable and nonlinear behavior of stock markets poses challenges for traditional forecasting models like ARIMA or GARCH. The authors review 148 studies that utilize AI techniques, including neural networks, support vector machines, and hybrid models. AI's ability to analyze large datasets and uncover hidden patterns makes it a promising tool for stock prediction. The review classifies the surveyed studies based on study characteristics (e.g., stock markets covered, input data) and model characteristics (e.g., data preprocessing, AI technique used). The study concludes that AI models, especially hybrid ones, have shown superior performance compared to traditional methods. However, the success of AI models depends on proper data selection, pre-processing, and training algorithms. The review highlights the growing interest in AI-based models since the 2008 financial crisis and the potential benefits for investors, corporations, and financial institutions if AI-based predictions are accurate.

2.1.7 Artificial Intelligence Applied to Stock Market Trading: A Review

[7] The paper reviews the application of Artificial Intelligence (AI) in stock market trading, analyzing 2,326 papers from Scopus between 1995 and 2019. It categorizes research into portfolio optimization, stock prediction, sentiment analysis, and combinations of these approaches, highlighting the increasing complexity and specificity in AI applications. The review underscores how AI has revolutionized stock market trading since the 1990s, with a significant growth in research in the AI-finance field. AI techniques have enhanced portfolio selection and risk management, while AI models are increasingly used to predict future stock prices based on historical data. Sentiment analysis, particularly analyzing news and social media sentiment, has become an important factor in market predictions.

A key finding is that the combination of multiple AI approaches improves trading outcomes, showcasing the importance of integrating different methods to achieve better results. The evolution of AI in finance has led to more complex models that incorporate various objectives and constraints, reflecting the sophisticated nature of financial markets. AI's versatility is evident across multiple areas, including portfolio management, price forecasting, and sentiment analysis, demonstrating its ability to tackle diverse financial challenges. Sentiment analysis, in particular, has emerged as a critical tool in predicting stock movements, emphasizing the role of investor psychology in market dynamics.

The paper also highlights the increasing need for robust AI models that can efficiently analyze and interpret vast amounts of complex financial data. Despite the advancements, there are still gaps in research, particularly in the comparative effectiveness of different feature sets in AI models. The authors suggest that continued research is necessary to optimize model performance. As AI technology continues to advance, it promises to further revolutionize financial markets, calling for ongoing exploration of new applications and methodologies in finance.

2.1.8 When AI Meets Finance (StockAgent): Large Language Model-based Stock Trading in Simulated Real-world Environments

[8] The paper introduces StockAgent, a cutting-edge multi-agent framework that employs large language models (LLMs) such as GPT and Gemini to simulate stock trading in a dynamic, event-driven environment. The purpose of StockAgent is to closely replicate real-world trading conditions, allowing for a more nuanced analysis of how external factors, including macroeconomic indicators and company fundamentals, influence trading behaviors and profitability. Traditional stock trading models often face challenges like test set leakage, where the model unintentionally uses information from the test set in the training phase, leading to overfitted and unreliable results. StockAgent overcomes this limitation by using LLMs to simulate real-time market conditions, capturing dynamic interactions between AI agents and external market forces.

The framework's design allows LLMs to function as autonomous agents, making decisions based on incoming data in a way that mimics human traders but is augmented by the expansive processing power of AI. This architecture enables StockAgent to explore various trading strategies under different market scenarios, shedding light on how AI models respond to specific events and environmental shifts. Unlike traditional models, which often rely on historical data alone, StockAgent's event-driven simulations enable a more robust analysis of how external influences, like changes in GDP or company earnings, impact stock prices in real-time.

One of the major contributions of StockAgent is its ability to highlight the differences in trading patterns between different LLMs. For instance, GPT may interpret market conditions differently than Gemini, leading to varied trading decisions and outcomes. This finding reveals that the choice of LLM can significantly affect the behavior of AI agents in the stock market, suggesting that certain models may be better suited to specific market conditions or types of assets. StockAgent also allows researchers to analyze the performance of different trading strategies, providing insights into the effectiveness of AI-driven investment approaches in diverse contexts.

Ultimately, this research contributes to the broader understanding of AI in finance, emphasizing the importance of accounting for external influences when developing AI-driven investment strategies. By simulating complex market dynamics and illustrating how AI agents respond to changing conditions, StockAgent offers valuable insights into the future of AI-driven trading. It highlights the potential for LLMs to play a critical role in improving investment decision-making, provided that the nuances of market behavior are carefully considered. However, the framework also acknowledges the need for further refinement to more accurately reflect the complexities of human behavior and market volatility, as these elements are crucial to developing truly reliable AI-based trading systems.

2.2 Limitation of Previous Studies

For the first paper, there are notable limitations when using Sentiment Analysis for stock market investment prediction. The accuracy of sentiment analysis in predicting stock market trends is heavily dependent on the quality and reliability of data sourced from social media platforms. The efficacy of this analysis relies on the veracity and representativeness of the sentiments captured, as well as the methodologies used to preprocess and analyze the data. Furthermore, the performance of artificial neural networks (ANN) in stock market prediction depends on the careful selection of parameters, such as the number of hidden layers and neurons, which requires extensive tuning to achieve optimal accuracy. However, relying solely on sentiment analysis may overlook other important factors like fundamental analysis and broader market trends, which also significantly influence stock prices. Additionally, ANN models' reliance on historical data can be problematic when market conditions and sentiments shift rapidly, rendering past data less indicative of future trends. Lastly, the effectiveness of sentiment analysis varies across companies and industries, necessitating a more nuanced approach to accommodate these differences when implementing predictive models.

The second research paper highlights several limitations, particularly in the time and cost associated with analyzing large datasets such as annual reports. The extensive nature of these reports can make the analysis process time-consuming, which may be impractical in fast-paced financial contexts. Moreover, the subjective evaluation of sections like the Business Overview, Risk Factors, and Management's Discussion and Analysis (MDA) poses challenges for large language models (LLMs), as these sections contain nuanced information that may be difficult for AI to accurately assess. Another limitation is the cost of querying LLMs for additional insights, especially when comprehensive contextual information is needed from the reports. These issues point to the need for further refinement of LLMs to improve their efficiency and effectiveness in analyzing complex financial documents.

The third paper presents several limitations, starting with its focus on the Japanese stock market, which limits its broader relevance to other markets and languages. It assumes that ChatGPT's sentiment analysis can accurately predict stock returns without fully addressing the uncertainties and limitations of this approach. Furthermore, the paper lacks a

comprehensive analysis of the performance measures, hindering a proper evaluation of the proposed investment strategy's reliability. Another issue is the difficulty in constructing an accurate polarity dictionary based on ChatGPT, particularly given the potential for language model biases. Finally, the paper does not adequately consider external factors and market conditions, which can greatly influence the effectiveness of sentiment-based strategies in stock market prediction.

The study on deep learning models titled "Which Artificial Intelligence Algorithm Better Predicts the Chinese Stock Market?" identifies several key limitations. One major issue is the model's data dependency—its effectiveness is heavily reliant on the quality and quantity of data, which can often be noisy or prone to errors in financial markets. Overfitting is another concern, as deep learning models trained on smaller datasets can capture noise rather than underlying patterns, leading to poor generalization on new data. Additionally, the computational complexity of deep learning models presents challenges, particularly for smaller firms or individual investors who may lack access to the necessary computational resources. The "black box" nature of these models also raises concerns about interpretability, as it can be difficult for investors to understand how predictions are made. Finally, the dynamic nature of financial markets means that models that perform well under specific conditions may not adapt effectively to changing market environments.

In the fifth paper, the authors acknowledge several limitations in their study. The prediction accuracy of around 70% is considered modest, and there is potential for improvement. They suggest that incorporating additional factors, such as market and company news, and using more powerful computational resources could enhance model performance.Another limitation is the sample size; the study used 82 randomly selected NYSE stocks, which may not be representative of the entire market. Including a broader range of stocks could lead to a more comprehensive analysis. The study relies solely on past price data and technical indicators, which, while improving results compared to previous research, might be enhanced by incorporating company-specific data. However, access to such data may be challenging for general traders and researchers.

The authors also noted their attempt to use neural networks was unsuccessful, despite suggestions from other sources. Future research could explore different neural network models and architectures.Finally, while the model's error rates for test and training sets were similar, indicating no overfitting, the study lacks details on evaluation metrics like precision, recall, or F1 score. Providing this information would be valuable for practical trading decisions. In summary, the study's limitations include a small sample size, reliance on past data, challenges with neural network implementation, and insufficient evaluation metrics. Addressing these issues could improve the model's accuracy and practical application.

The sixth paper, "Application of Artificial Intelligence in Stock Market Forecasting: A Critique, Review, and Research Agenda," points out several limitations. First, the quality of input data is crucial—insufficient or poorly selected data can lead to flawed predictions. Many studies reviewed in the paper fail to adequately select or preprocess input variables, compromising model performance. AI models, especially neural networks, face computational challenges when processing large datasets, which can limit their scalability. Overfitting is also a concern, particularly when models are trained on noisy or limited data. Another significant limitation is the lack of interpretability in AI models, which are often seen as opaque or “black boxes.” Finally, the fast-changing nature of financial markets means that models performing well in certain conditions may not adapt effectively to different economic environments or regulatory changes.

The seventh paper on the application of AI in stock market trading highlights several limitations, including the reliance on historical data, which may not accurately reflect future market behavior. Many AI models are also highly complex and opaque, making it difficult for users to understand how predictions are made. Additionally, external factors such as geopolitical events and economic shifts, which significantly influence market dynamics, may not be sufficiently accounted for in AI models. This can lead to overfitting, where models perform well on historical data but fail to generalize in real-world scenarios. High computational costs and the need for large amounts of data can also limit accessibility for smaller investors or firms.

Finally, the paper "When AI Meets Finance (StockAgent): Large Language Model-based Stock Trading in Simulated Real-world Environments" identifies several key limitations. A major issue is the reliance on large language models (LLMs) such as GPT and Gemini, which can introduce biases based on the data they are trained on, leading to skewed trading recommendations. LLMs are also not specifically designed for stock trading and may lack the necessary financial market expertise to model certain trading strategies accurately. Another significant limitation is StockAgent's inability to fully capture the complexities of human emotions and irrational behaviors, which are key drivers in real-world trading environments. The framework also assumes ideal market conditions, which may oversimplify trading during periods of extreme volatility or crises. Additionally, StockAgent's simulations are limited in scope, focusing on specific market scenarios, which reduces its generalizability to other markets or asset classes.

2.3 Proposed Solutions

To address the weaknesses and limitations, researchers and developers are actively working on improving the transparency and interpretability of AI models in financial applications. Efforts are underway to develop AI systems that can explain their decision-making processes, providing users with more confidence and trust. Additionally, advancements in data quality and preprocessing techniques are helping mitigate the impact of noisy or biased data on AI model performance. To make AI-driven financial tools more accessible, there is a push for the development of cloud-based solutions and APIs that can be readily integrated into various financial platforms. These innovations are crucial for expanding the usability of AI technologies across different sectors of finance, making them more scalable and versatile for diverse applications.

To overcome the limitations of the second paper, several strategies can be employed. Incorporating more recent financial news data alongside historical datasets can enhance the model's adaptability to evolving market conditions, ensuring more dynamic responses to real-time events. Robustness testing should encompass various market scenarios, including high volatility and economic crises, to gauge the model's performance under diverse conditions. Diversifying the dataset by including news articles from a broader array of sources can mitigate source bias and improve prediction accuracy. Furthermore, evaluating the profitability of the trading strategy should account for trading costs like transaction fees for a more realistic assessment. Expanding the scope to include different model architectures, intraday prediction timeframes, and real-world trading analyses will provide a more comprehensive evaluation of the model's effectiveness. Extending the prediction horizon for long-term forecasting is also essential for gauging sustained model performance across different timeframes.

To tackle the limitations identified in the third paper, future research endeavors can extend the scope of sentiment analysis by ChatGPT beyond the Japanese stock market. This expansion would involve exploring the applicability of ChatGPT's sentiment analysis capabilities to a broader array of markets and languages, enhancing the generalizability of findings. Additionally, addressing the assumption of unwavering accuracy in sentiment analysis based on ChatGPT requires further investigations to meticulously evaluate the

potential limitations and uncertainties inherent to this approach. These assessments should examine how model biases or constraints impact predictions, providing a more nuanced understanding of the reliability of sentiment analysis in stock market prediction. To bolster the robustness of the proposed investment strategy, future studies should analyze performance measures in greater detail, emphasizing transparency and reproducibility. This approach will ensure a more rigorous evaluation of the strategy's effectiveness and address challenges in constructing a polarity dictionary using ChatGPT.

In addressing the limitations presented in the fourth paper, several solutions can be implemented. Data preprocessing and augmentation techniques can be employed to clean and normalize data, thereby reducing noise and improving model reliability. Augmenting the training dataset can help mitigate overfitting by creating a more diverse data environment. Regularization techniques, such as dropout, L1/L2 regularization, or early stopping, can further combat overfitting by ensuring that the model generalizes better to unseen data. Transfer learning offers another solution, allowing a model trained on a large, generalized dataset to be fine-tuned on a smaller, domain-specific dataset, thus enhancing performance while minimizing computational demands. Incorporating explainability tools such as SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations) can improve transparency by elucidating the model’s decision-making process. Developing adaptive learning algorithms that adjust to real-time changes in the financial environment can help maintain predictive accuracy, enabling models to continuously learn from new data.

To address the limitations in “Using AI to Make Predictions on the Stock Market,” several strategies are proposed. The authors propose several solutions to enhance the accuracy of short-term stock price prediction. They recommend incorporating additional factors, such as news about the stock market and specific companies, which can significantly influence stock prices. Leveraging natural language processing to analyze news articles and social media sentiment could offer valuable insights. The authors also suggest using machines with higher computational power to process larger datasets and apply more complex algorithms, potentially uncovering more intricate patterns and improving prediction accuracy. Furthermore, they propose exploring different neural network models, despite their initial attempts being unsuccessful, as other research indicates their potential utility. To provide a

more comprehensive evaluation of the model's performance, the authors highlight the need to include detailed metrics like precision, recall, and F1 score. Lastly, they acknowledge the limitation of their sample size and advocate for including a more diverse set of stocks from various industries and markets to achieve a more representative analysis. These improvements aim to enhance the model's predictive power and practical applicability in trading scenarios.

For the paper "Application of Artificial Intelligence in Stock Market Forecasting: A Critique, Review, and Research Agenda," several solutions are proposed to enhance AI model performance. Rigorous data preprocessing techniques, such as normalization and dimensionality reduction, should be applied to improve model accuracy and reduce noise. Data augmentation strategies can further enhance the diversity of training datasets, mitigating overfitting risks. The use of hybrid models combining AI techniques, such as fuzzy logic with neural networks, is recommended for improved prediction accuracy. Tools like SHAP or LIME should be integrated to make AI models more interpretable and transparent, which can help investors understand how decisions are made. Lastly, adaptive models that continuously learn from new data and adjust to real-time market conditions are crucial for maintaining predictive accuracy, especially as market dynamics evolve and new financial trends emerge.

To overcome the limitations of AI in stock market trading presented in the seventh paper, researchers can focus on developing more interpretable models that allow users to understand how predictions are made. Incorporating diverse data sources, including alternative data like social media sentiment, can enhance model robustness and provide a fuller picture of market sentiment. Adopting ensemble methods that combine multiple AI models may help mitigate overfitting while improving prediction accuracy. Continuous model training with real-time data will ensure that models can adapt to evolving market dynamics, while collaborations between financial experts and data scientists can lead to the development of hybrid models that blend traditional financial theories with advanced AI techniques for better predictions.

Finally, to address the limitations of StockAgent, several enhancements are proposed to improve its effectiveness and reliability. One primary recommendation is diversifying the range of LLMs used in the framework. By incorporating a broader selection of models

trained on diverse datasets, StockAgent can reduce bias and improve the reliability of its trading recommendations. Incorporating elements from behavioral finance into StockAgent would allow the framework to simulate more realistic trading behaviors, accounting for emotions like fear and greed. Expanding the range of market conditions simulated—including extreme volatility and financial crises—would improve the system's robustness. Additionally, integrating real-time data feeds and adaptive learning mechanisms would enable StockAgent to respond dynamically to changing market conditions, refining its trading strategies. Finally, developing interpretability tools such as SHAP or LIME would improve transparency, helping users understand the rationale behind AI-driven trading decisions, thereby building confidence in AI-powered investment strategies.

CHAPTER 3

System Methodology/Approach

3.1 System Design Diagram

3.1.1 System Architecture Diagram

This project employs a data-driven approach to analyze sentiment in YouTube videos and news articles related to the stock market, aiming to extract valuable insights and provide recommendations for informed investment decisions. The systems for YouTube and news analysis operate independently but follow similar principles.

For YouTube videos, the project begins by utilizing the YouTube Data API v3 to identify the latest video IDs from user-specified channels. This focused approach ensures that only relevant and recent stock market discussions are analyzed. Once video IDs are obtained, transcripts are retrieved using the YouTube Transcript API. These transcripts are then preprocessed, which involves cleaning and formatting the text to create a unified block suitable for sentiment analysis. OpenAI's API, equipped with a fine-tuned model, analyzes the sentiment of the preprocessed transcripts within the stock market context. The analysis generates a detailed summary, which includes overall sentiment, sentiment towards individual stocks, actionable investment insights, potential sentiment shifts, and a sentiment score. Additionally, the model may provide stock market predictions based on the sentiment analysis. To present the findings professionally, reportlab is used to convert the analysis results into formatted PDF documents.

In parallel, the news article analysis system employs Beautiful Soup 4 (bs4) to scrape content from financial websites. Users can provide specific URLs or perform keyword-based searches to locate relevant articles. Beautiful Soup parses the retrieved webpages to extract the necessary text. This text is then preprocessed similarly to the YouTube transcripts, creating a coherent block of text for sentiment analysis. OpenAI's API performs the sentiment analysis, generating a comprehensive summary that includes overall sentiment, sentiment towards individual stocks, actionable investment insights, potential sentiment shifts, and a

sentiment score. An optional step can include relationship analysis to explore connections between sentiment and stock price movements. The analysis results are also converted into professionally formatted PDFs using reportlab, ensuring that the findings are presented in a clear and organized manner.

Throughout both systems, error handling mechanisms are incorporated to manage potential issues such as failed data retrieval, API errors, and missing data, ensuring a robust and reliable analysis process.

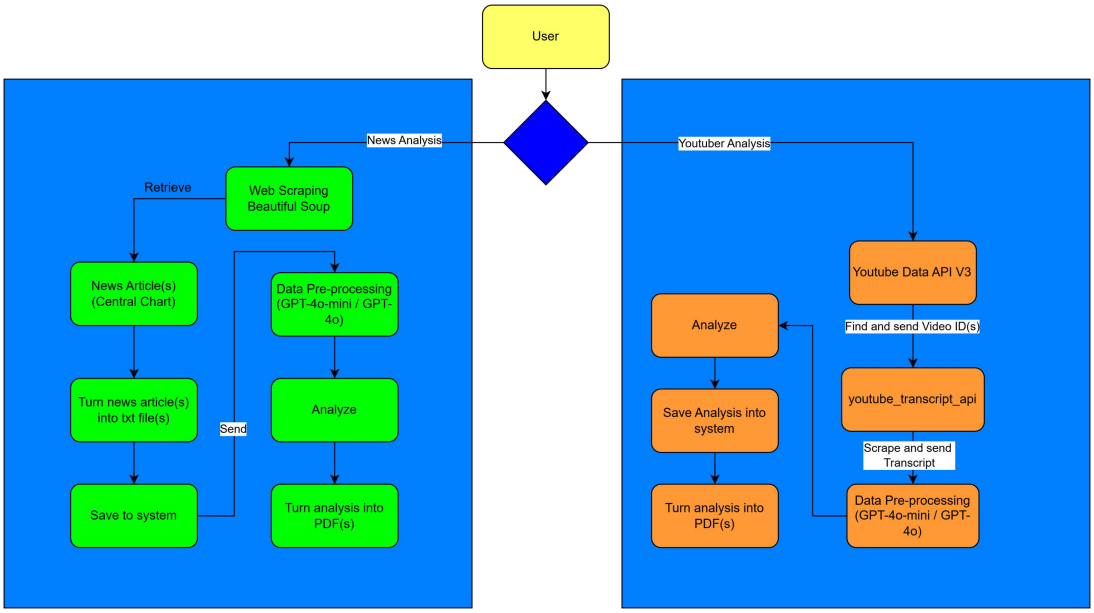


Figure 3.1 General Work Flow of System

3.1.2 Use Case Diagram and Description

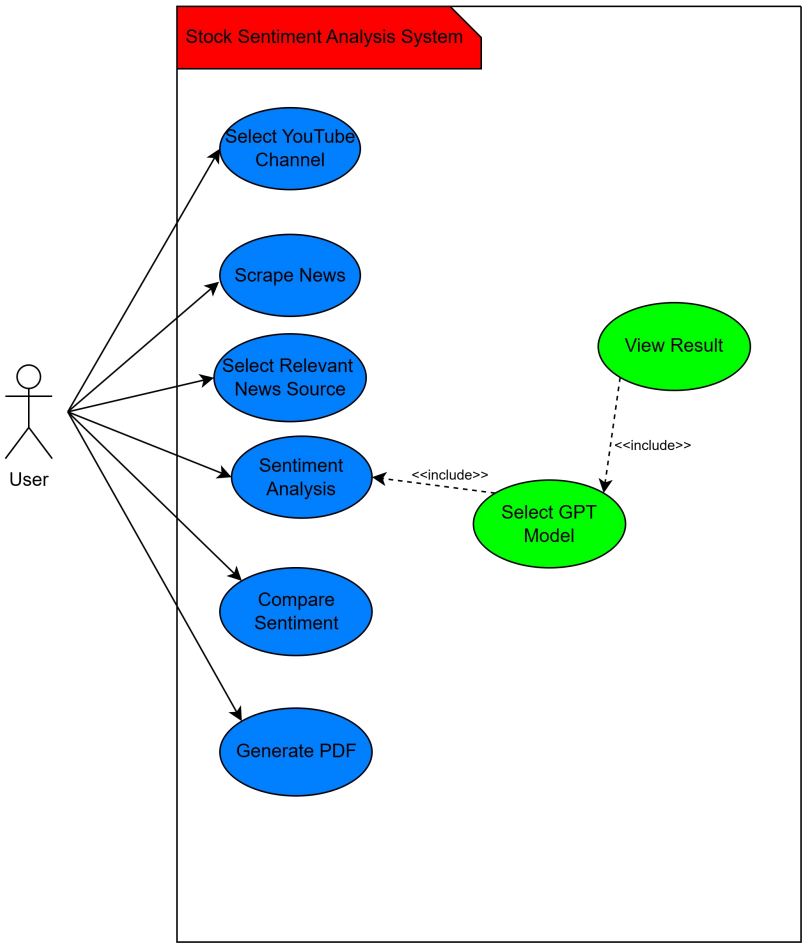


Figure 3.2 Use Case Diagram for the System

The use case diagram illustrates the interaction between the User and the Stock Sentiment Analysis System. The system allows the user to conduct a comprehensive analysis of stock sentiment by integrating multiple sources such as YouTube channels and news articles. The process begins with the user selecting a YouTube channel of interest, after which the system scrapes relevant news articles. Once the relevant news sources are identified, the user can initiate the sentiment analysis, which the system carries out based on the data gathered from both YouTube and news sources. The user is also given the option to select the GPT model that will be used for sentiment analysis, ensuring flexibility in the depth and accuracy of the results. Once the sentiment analysis is completed, the user can view the results. Additionally, the system provides a feature to compare sentiments from different sources, allowing the user to make informed decisions based on the synthesized data. Finally, the user has the ability to

generate a PDF that compiles the sentiment analysis and comparisons, enabling them to save or share the results. The system supports the entire workflow of sentiment analysis, from gathering data to producing a final report in a convenient PDF format.

3.1.3 Activity Diagram

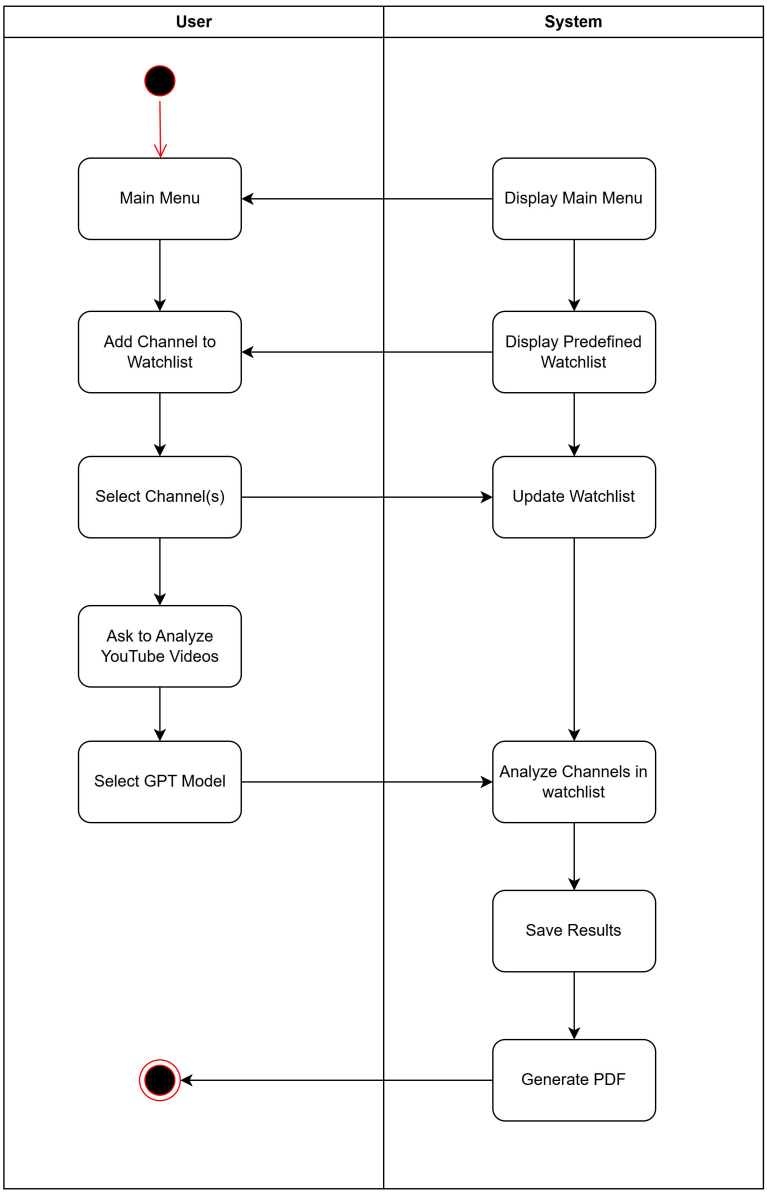


Figure 3.3 Activity Diagram for the YouTuber Sentiment Analysis System

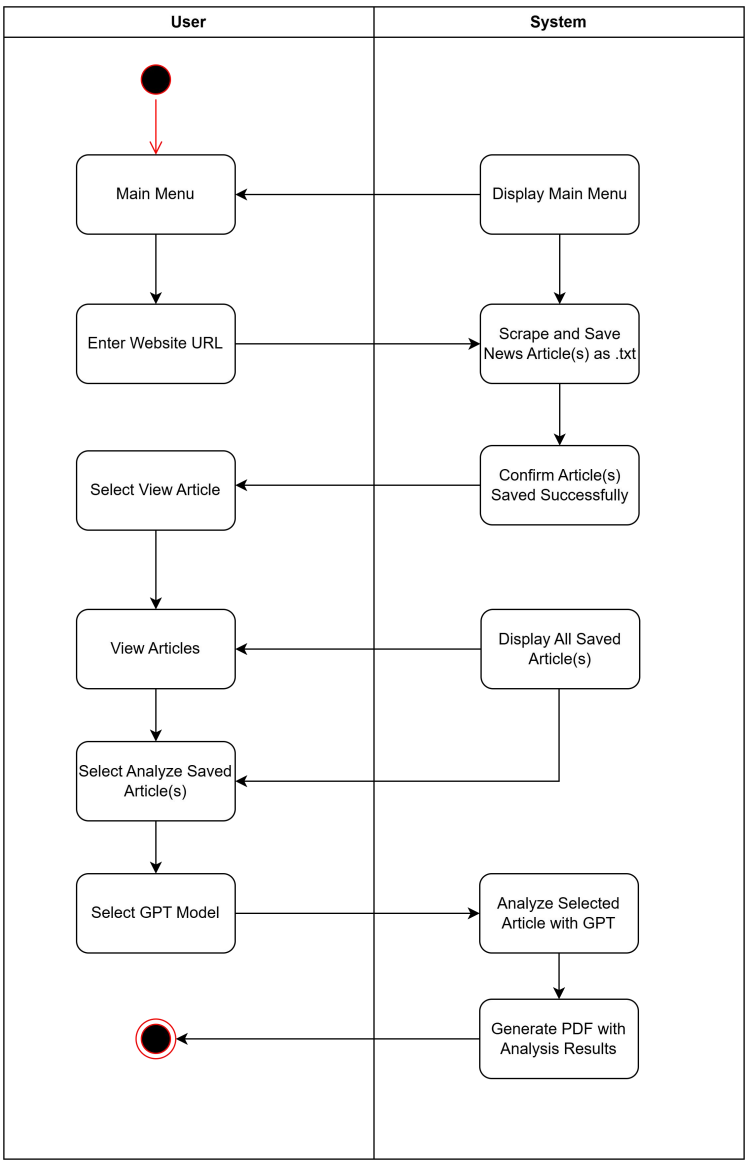


Figure 3.4 Activity Diagram for the News Sentiment Analysis System

Chapter 4

System Design

4.1 System Block Diagram (YouTube)

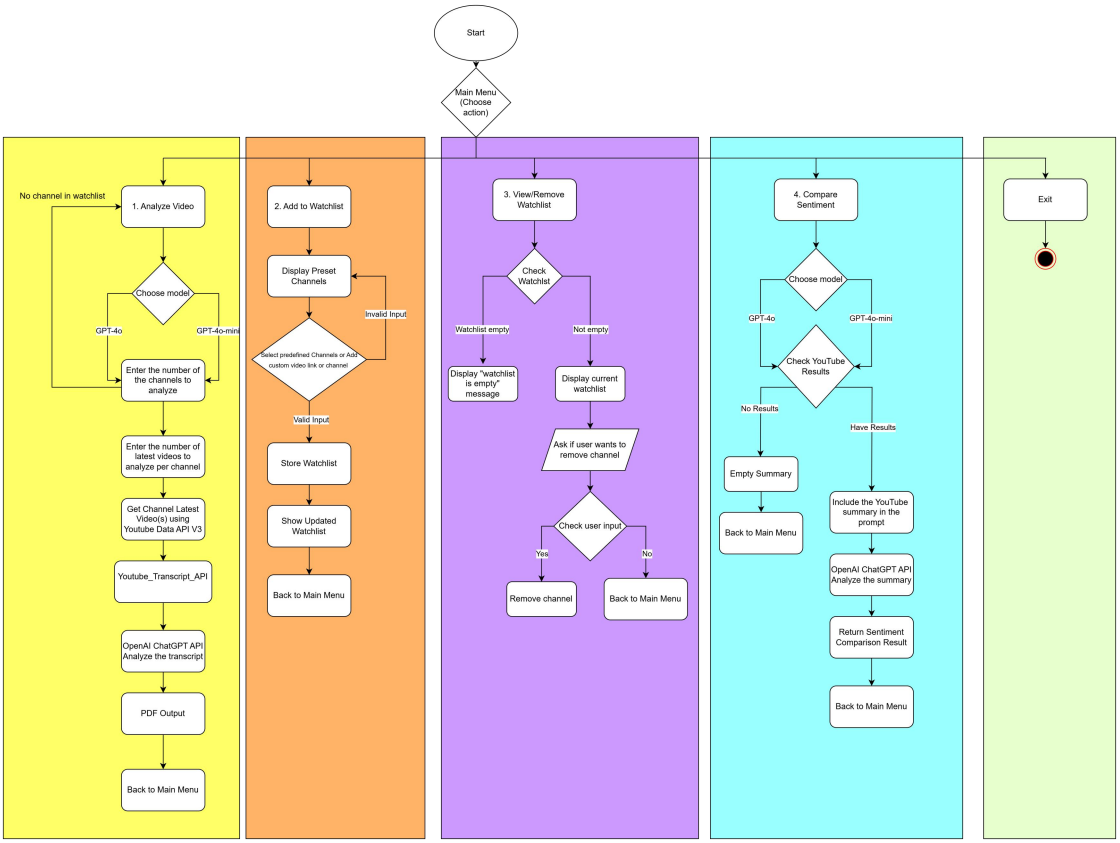


Figure 4.1 General Work Procedure of YouTuber Sentiment Analysis System

4.1.1 Youtube Data API V3

The purpose of Youtube Data Api V3 is to automatically collect and gather information of the selected youtube channel. There are two sources of information that the system will get for now: Youtuber's latest video, the video's comments. Depending on the user's choice of action, the API will either pass to the youtube_transcript_api or straight to the Model(OpenAI API). In the future more data will be collected to provide more information to the GPT model, for it to analyze as well as achieve higher accuracy prediction.

4.1.2 Youtube_Transcript_API

The purpose of youtube_transcript_api lies in extracting valuable text data from YouTube videos. This text data, typically in the form of transcripts, serves as a crucial information source for the system. Similar to the Youtube Data API v3, the initial focus is on the latest video and its corresponding transcript. Depending on the user's needs, the extracted transcript can be either fed directly to the OpenAI model or passed on for further processing. In the future, the API might be expanded to handle transcripts from a broader range of videos, providing the model with a richer dataset for analysis and potentially leading to more accurate predictions.

4.1.3 ReportLab

ReportLab is a robust Python library used for generating PDFs programmatically. It allows for the creation of well-structured, detailed PDF reports with customized layouts, fonts, and formatting options, including text wrapping, margins, and the integration of images such as logos. The library supports the merging of individual PDF files into one comprehensive document, making it suitable for automating the creation of polished reports from analyzed data. Its flexibility and ease of use make it ideal for generating professional-quality PDFs in various applications.

4.1.4 Data Pre-processing

In this project, tasks were distributed across multiple GPT models to efficiently preprocess the data before it was used for sentiment analysis. By segmenting the tasks across different models, we minimized the number of input tokens required and lowered costs associated with API calls. Four distinct GPT models were assigned to handle separate tasks as follows:

i. YouTube Video Summarization

For each selected YouTube video related to stock market trends, the GPT-4o API was used to generate detailed summaries of the content. The API focused on key sections of the video, including the host's sentiment, main stock market predictions, investment advice, and

discussion of industry trends. The designated GPT-4o model extracted these core insights and formatted the output into structured sections like Overall Sentiment, Investment Outlook, Key Stocks Mentioned, and Market Predictions. The results were then saved into JSON format for further analysis and report generation, ensuring that the data could be efficiently stored and processed.

ii. Sentiment Comparison Analysis

A separate GPT model was used to compare sentiments between YouTube video summaries and articles. The GPT-4o API was tasked with analyzing differences in investment advice, opinions, and outlooks across multiple sources. By feeding in preprocessed summaries of YouTube videos and news articles, the model generated comparative insights and identified key sentiment shifts between the sources. The output of this task included overall sentiment trends, actionability of insights, and significant opinion contrasts, which were used to guide further investment decisions.

iii. Stock Market Report Generation

The GPT-4o API was also employed to generate comprehensive reports based on the results from YouTube video and news analysis. The reports provided investment advice, sentiment scores, and actionable insights. This report generation task focused on consolidating data from multiple sources, allowing the model to analyze cross-referenced trends and build a final assessment of stock market conditions. The generated reports followed a structured format, making it easier for users to understand the insights and predictions. These reports were later converted into PDFs for record-keeping and further usage.

iv. Trend and Credibility Analysis

Instead of relying solely on single-source data, the system also utilized GPT to conduct trend analysis based on ongoing market news and stock reports. This analysis evaluated the credibility of investment advice and predictions from various sources over time, checking consistency between past and current statements. The model tracked historical predictions against actual market trends to assess the credibility of key market influencers. By integrating this with external news reports and company financials, the system could better gauge market

sentiment and predict future directions with higher accuracy. The output provided a credibility rating along with qualitative analysis of the influencers' predictions.

4.1.5 Model Evaluation for YouTuber Sentiment Analysis System

In our model evaluation approach, we assess the qualitative data generated by the system, with a particular emphasis on the recommendations and justifications it provides to users. The goal is to ensure that the information conveyed is accurate and fact-based. In later sections, we carried out tests on three YouTubers and across ten different companies.

4.2 System Block Diagram (News)

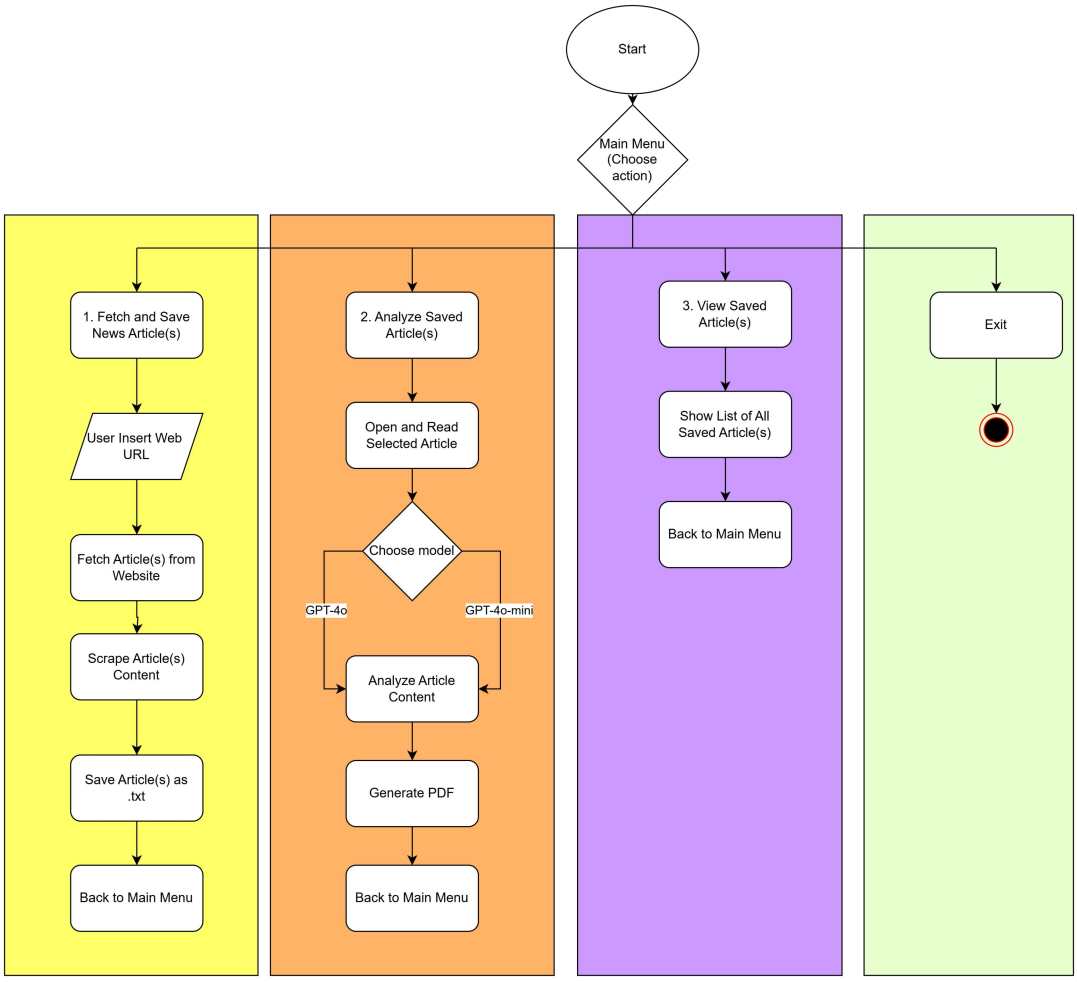


Figure 4.2 General Work Procedure of News Sentiment Analysis System

4.2.1 Beautiful Soup 4

Beautiful Soup 4 acts as web scraping workhorse, tasked with gathering relevant information from financial websites. Similar to the YouTube Data API v3, it focuses on user-specified URLs or crawls based on keywords related to the stock market. Unlike video data, Beautiful Soup extracts the website's text content, like news articles. This text is then fed into the system for further analysis by the OpenAI model. Other than that, the text is also passed into a separate html file to enhance the user's reading experience. As the project evolves, Beautiful Soup 4 will be used to scrape data from a wider range of financial sources, providing the model with a more comprehensive dataset for analysis and potentially leading to more accurate insights.

4.2.2 ReportLab

ReportLab is a robust Python library used for generating PDFs programmatically. It allows for the creation of well-structured, detailed PDF reports with customized layouts, fonts, and formatting options, including text wrapping, margins, and the integration of images such as logos. The library supports the merging of individual PDF files into one comprehensive document, making it suitable for automating the creation of polished reports from analyzed data. Its flexibility and ease of use make it ideal for generating professional-quality PDFs in various applications.

4.2.3 Data Pre-processing

In this project, tasks were distributed across multiple GPT models to efficiently preprocess the data before it was used for sentiment analysis. By segmenting the tasks across different models, we minimized the number of input tokens required and lowered costs associated with API calls. Four distinct GPT models were assigned to handle separate tasks as follows:

i. YouTube Video Summarization

For each selected YouTube video related to stock market trends, the GPT-4o API was used to generate detailed summaries of the content. The API focused on key sections of the video,

including the host's sentiment, main stock market predictions, investment advice, and discussion of industry trends. The designated GPT-4o model extracted these core insights and formatted the output into structured sections like Overall Sentiment, Investment Outlook, Key Stocks Mentioned, and Market Predictions. The results were then saved into JSON format for further analysis and report generation, ensuring that the data could be efficiently stored and processed.

ii. Sentiment Comparison Analysis

A separate GPT model was used to compare sentiments between YouTube video summaries and articles. The GPT-4o API was tasked with analyzing differences in investment advice, opinions, and outlooks across multiple sources. By feeding in preprocessed summaries of YouTube videos and news articles, the model generated comparative insights and identified key sentiment shifts between the sources. The output of this task included overall sentiment trends, actionability of insights, and significant opinion contrasts, which were used to guide further investment decisions.

iii. Stock Market Report Generation

The GPT-4o API was also employed to generate comprehensive reports based on the results from YouTube video and news analysis. The reports provided investment advice, sentiment scores, and actionable insights. This report generation task focused on consolidating data from multiple sources, allowing the model to analyze cross-referenced trends and build a final assessment of stock market conditions. The generated reports followed a structured format, making it easier for users to understand the insights and predictions. These reports were later converted into PDFs for record-keeping and further usage.

iv. Trend and Credibility Analysis

Instead of relying solely on single-source data, the system also utilized GPT to conduct trend analysis based on ongoing market news and stock reports. This analysis evaluated the credibility of investment advice and predictions from various sources over time, checking consistency between past and current statements. The model tracked historical predictions against actual market trends to assess the credibility of key market influencers. By integrating

this with external news reports and company financials, the system could better gauge market sentiment and predict future directions with higher accuracy. The output provided a credibility rating along with qualitative analysis of the influencers' predictions.

4.2.4 Model Evaluation for News Sentiment Analysis System

In our model evaluation approach for the news system, we focus on assessing the qualitative data it produces, particularly the recommendations and justifications offered to users. This is done to ensure that the information provided is accurate and based on factual evidence. In the following sections, we conducted tests analyzing five companies using news articles.

Chapter 5

System Implementation

5.1 Hardward Setup

Description	Specifications
Model	Laptop
Processor	AMD Ryzen 7 4800H with Radeon Graphics
Operating System	Windows 11
Graphic Card	NVIDIA GeForce GTX 1650Ti 4GB
Memory	8GB RAM
Storage	512GB NVME M.2 SSD

Table 5.1 Specifications of laptop

5.2 Software Setup

1. Beautiful Soup 4

This Python library acts like a soup ladle for navigating websites. It parses HTML content into a structured format, allowing you to easily search and extract specific data like text, headings, or links from webpages. This makes it a valuable tool for tasks like data scraping and content analysis.

2. Youtube_Transcript_Api

This library bridges the gap between your Python code and YouTube transcripts. It acts as a messenger, retrieving the transcript text for a specified YouTube video ID through interaction with external services. This enables you to analyze the content of YouTube videos programmatically, opening doors for tasks like sentiment analysis or summarizing video discussions.

3. YouTube Data Api V3

This official API from Google unlocks the power of YouTube data within your Python applications. By providing a developer key for authorization, you can leverage various functionalities like searching for videos based on keywords, retrieving detailed information about specific videos (title, description, view count), or managing playlists. This API empowers you to automate tasks and build applications that interact with YouTube data programmatically.

4. PyCharm

PyCharm is a powerful IDE from JetBrains that provides a comprehensive environment for writing, debugging, and testing Python code. It offers features like syntax highlighting, code completion, code refactoring, a built-in debugger, and integration with various version control systems. These features significantly enhance developer productivity and streamline the Python development process.

5. OpenAI

OpenAI offers access to powerful large language models (LLMs) like "gpt-3.5-turbo-0125" through their API. These models can perform various tasks like text generation, translation, and code completion. The code you provided utilizes the OpenAI library, which acts as a client-side interface, allowing you to interact with OpenAI's API and leverage their LLMs for tasks like sentiment analysis on the provided text data.

6. ReportLab

ReportLab is a robust Python library used for generating PDFs programmatically. It allows for the creation of well-structured, detailed PDF reports with customized layouts, fonts, and formatting options, including text wrapping, margins, and the integration of images such as logos. The library supports the merging of individual PDF files into one comprehensive document, making it suitable for automating the creation of polished reports from analyzed.

Particulars	Tools
Operating System	Microsoft Windows 11
Integrated Development Environment	PyCharm
Programming Language	Python
Libraries	BeautifulSoup4, OpenAI, youtube_transcript_api, YouTube Data Api V3, Reportlab

Table 5.2 Software Tools for Development

5.3 Setting and Configuration

The settings and configuration of the system are designed to provide flexibility, scalability, and enhanced user control across its various modules. The system, which focuses on analyzing YouTube videos and news articles for comparative sentiment analysis and generating PDF reports, offers dynamic options for customizing workflows, managing costs, and adjusting the depth of analysis. One of the key configuration options is the selection of the appropriate GPT API model for analysis. Users can choose between models such as GPT-4o-mini, which provides cost-effective, faster analysis with fewer tokens, ideal for high-volume, basic tasks, and GPT-4-turbo, which offers more complex and nuanced analysis at a higher cost. The system also allows users to configure automatic switching between models based on data complexity or budgetary limits, ensuring an optimal balance between cost efficiency and output quality.

In terms of watchlist management, users have the ability to customize their watchlist by selecting specific YouTube channels and news sources to monitor for relevant content. This includes predefined watchlists with popular sources for quick setup, as well as custom watchlists for manually added, niche, or specialized data.

5.4 System Operation (with Screenshots)

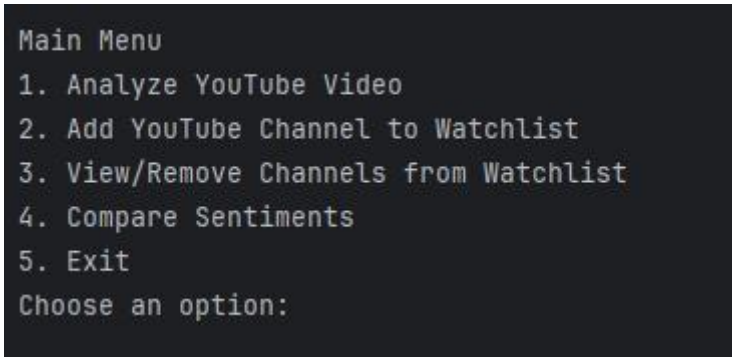


Figure 5.4.1 Main Menu for Youtube_Main (PyCharm Terminal)

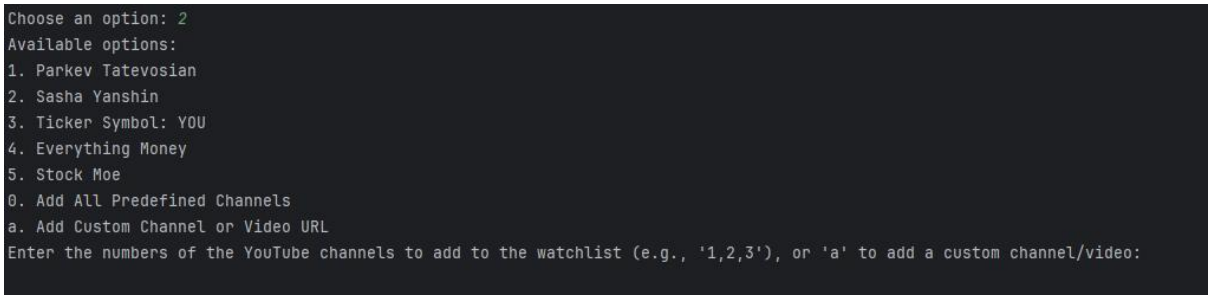


Figure 5.4.2 Add YouTube Channel To Watchlist (PyCharm Terminal)

Figure 5.4.3 Updated Watchlist and Confirmation on Watchlist (PyCharm Terminal)

```
Choose an option: 3
Current Watchlist:
1. Parkev Tatevosian
2. Ticker Symbol: YOU
Do you want to remove a channel from the watchlist? (y/n): y
Select the channel number to remove from the watchlist:
1. Parkev Tatevosian
2. Ticker Symbol: YOU
Enter the number(s) of the channel(s) to remove (e.g., '1,3'): 1
Removed 'Parkev Tatevosian' from the watchlist.
Updated Watchlist:
Current Watchlist:
1. Ticker Symbol: YOU
Do you want to remove a channel from the watchlist? (y/n): n
```

Figure 5.4.4 View/Remove Channels from Watchlist and Confirmation on Watchlist (PyCharm Terminal)

```
Choose an option: 1
Available models:
1. GPT-4o-mini - Cheaper
2. GPT-4-o - Smartest
Choose a model (1 or 2): 1
Current Watchlist:
1. Ticker Symbol: YOU
Enter the numbers of the channels to analyze (e.g., '1,3'): 1
Enter the number of latest videos to analyze per channel: 1
Analyzing video (ID: DEmwOcKiUU8) from channel 'Ticker Symbol: YOU'
Analysis complete. Generating PDF...
PDF generated and saved at: C:\Users\User\PycharmProjects\FYP1\Youtube_Main\PDFs\Ticker Symbol YOU_DEmwOcKiUU8.pdf
```

Figure 5.4.5 Analyze YouTube Video (PyCharm Terminal)

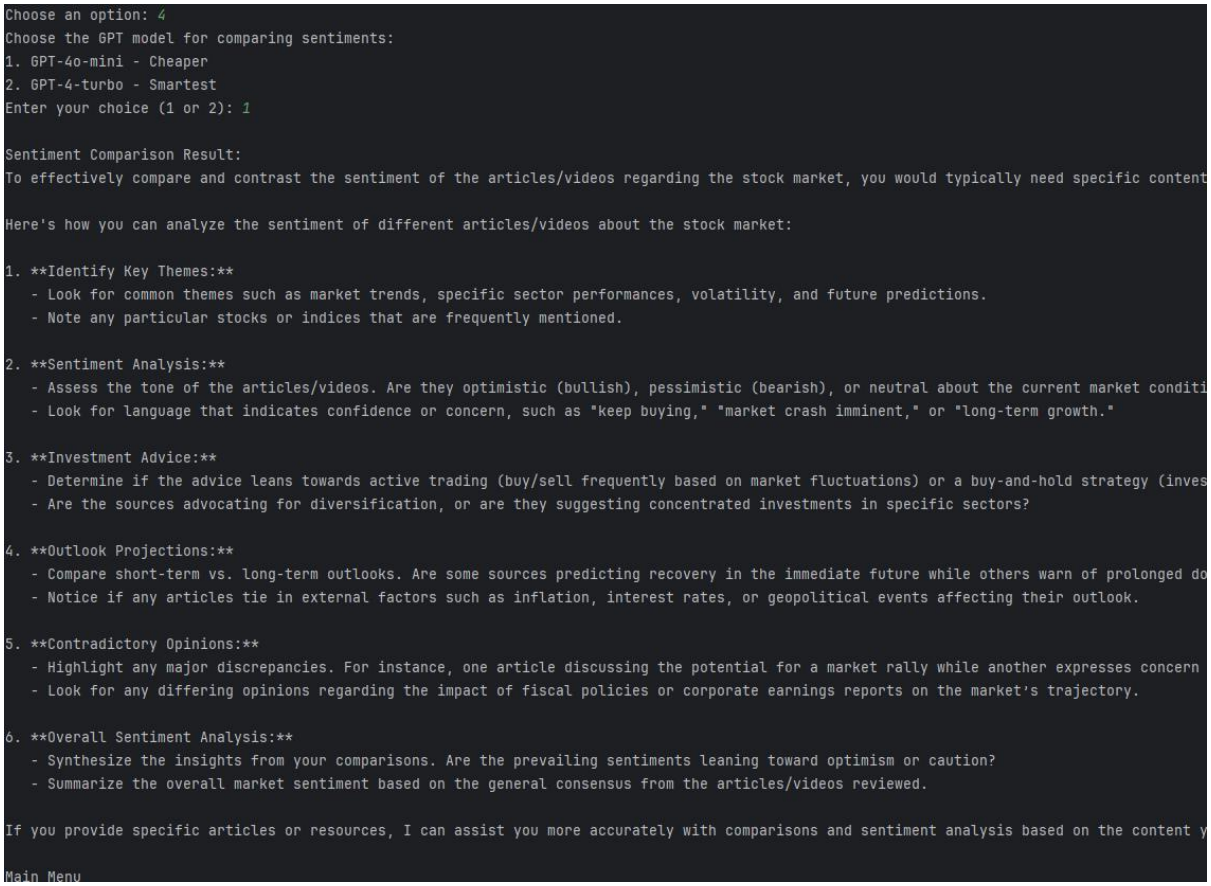


Figure 5.4.6 Compare Sentiment (PyCharm Terminal)

Initial idea was to present compared sentiment on webpage for user, to enable a dynamic view (comparison depend on their past analyzed videos/news). Due to an unfortunate turn of event the webpage (User Interface) was forced to dropped from this project. Hence, the result will be presented in the PyCharm Terminal.

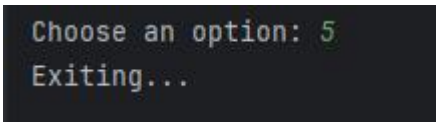
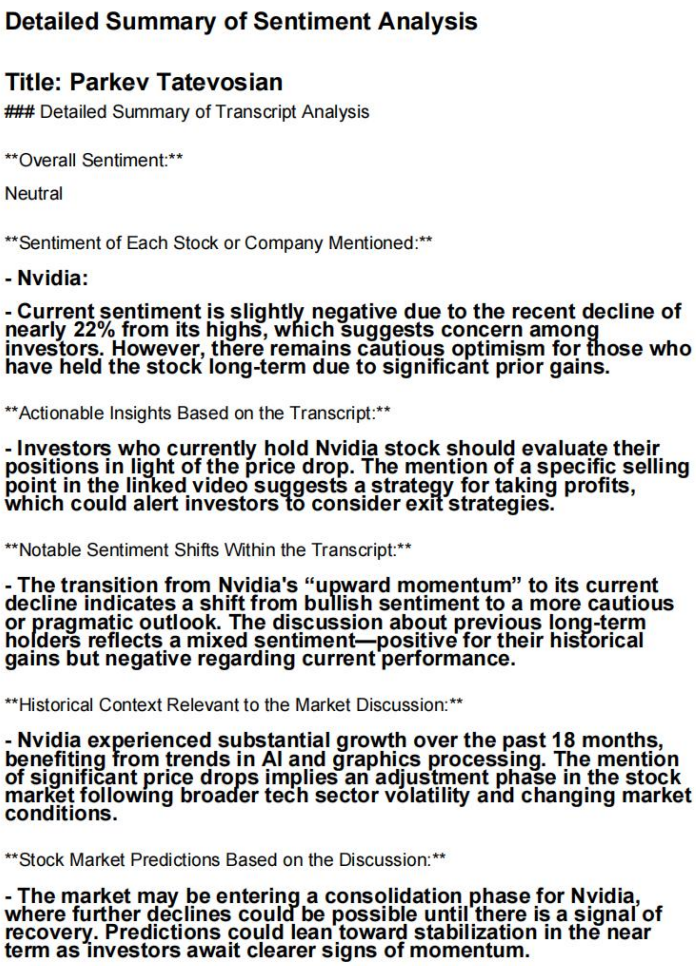


Figure 5.4.7 Exit (PyCharm Terminal)



Detailed Summary of Sentiment Analysis

Title: Parkev Tatevosian

Detailed Summary of Transcript Analysis

Overall Sentiment:

Neutral

Sentiment of Each Stock or Company Mentioned:

- Nvidia:

- Current sentiment is slightly negative due to the recent decline of nearly 22% from its highs, which suggests concern among investors. However, there remains cautious optimism for those who have held the stock long-term due to significant prior gains.

Actionable Insights Based on the Transcript:

- Investors who currently hold Nvidia stock should evaluate their positions in light of the price drop. The mention of a specific selling point in the linked video suggests a strategy for taking profits, which could alert investors to consider exit strategies.

Notable Sentiment Shifts Within the Transcript:

- The transition from Nvidia's "upward momentum" to its current decline indicates a shift from bullish sentiment to a more cautious or pragmatic outlook. The discussion about previous long-term holders reflects a mixed sentiment—positive for their historical gains but negative regarding current performance.

Historical Context Relevant to the Market Discussion:

- Nvidia experienced substantial growth over the past 18 months, benefiting from trends in AI and graphics processing. The mention of significant price drops implies an adjustment phase in the stock market following broader tech sector volatility and changing market conditions.

Stock Market Predictions Based on the Discussion:

- The market may be entering a consolidation phase for Nvidia, where further declines could be possible until there is a signal of recovery. Predictions could lean toward stabilization in the near term as investors await clearer signs of momentum.

Figure 5.4.8 Example output of generated PDF for YouTuber Sentiment Analysis

The YouTube-generated PDFs offer in-depth sentiment analysis of video transcripts from selected YouTubers. These reports summarize the overall sentiment towards companies, historical context, and investment advice, tailored to the unique perspectives presented in the videos.

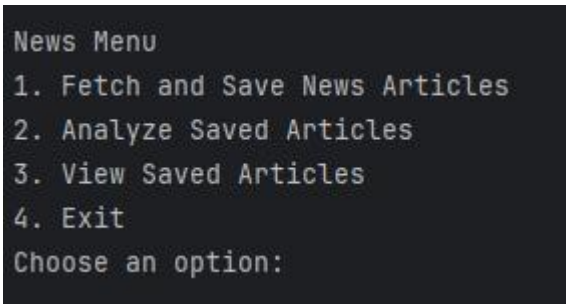


Figure 5.4.9 Main Menu for News_Main (PyCharm Terminal)

Figure 5.4.10 Fetch and Save News Articles (PyCharm Terminal)

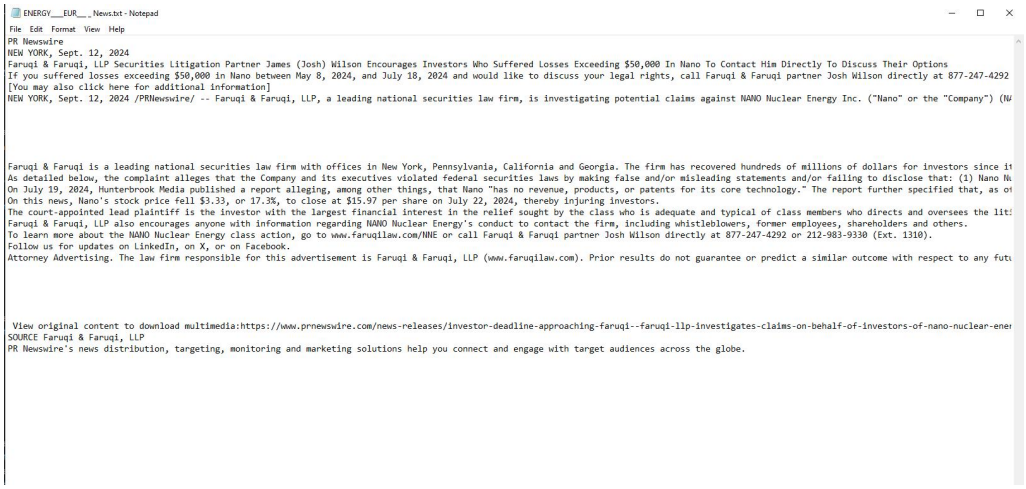


Figure 5.4.11 Example output of scarped news turned into .txt

```
Choose an option: 3

Saved Articles:
1. ALTECH ADV_MAT_ NA O_N_____EUR___ _ News.txt
2. AMERICAN AIRLINES GRP____EUR___ _ News.txt
3. BLACKROCK INC_____USD___ _ News.txt
4. BOSSARD N____CHF___ _ News.txt
5. BURKHALTER N____CHF___ _ News.txt
6. CENTENE CORP_____USD___ _ News.txt
7. CRITICAL RESOURCES LIMITED____AUD___ _ News.txt
8. DEXUS____AUD___ _ News.txt
9. DNO ASA A NK __25 _ News.txt
10. DOWNING ONE VCT ORD 1P _ News.txt
11. ENERGY____EUR___ _ News.txt
12. FALCON OIL _ GAS LTD_ COM SHS NPV _DI_____GBX___ _ News.txt
13. FANGDD NETWORK GROUP____USD___ _ News.txt
14. GLOBAL WATER RESOURCES INC_____USD___ _ News.txt
15. ING GROEP N_V_____EUR___ _ News.txt
16. INVESTIS N____CHF___ _ News.txt
17. INVESTOR AB _CBOE_____SEK___ _ News.txt
18. LATIN METALS INC_ LMSQF____USD___ _ News.txt
19. MISTER SPEX SE INH O_N_____EUR___ _ News.txt
20. News.txt
21. NOKIA____EUR___ _ News.txt
22. OMA SAASTOPANKKI OYJ _CBOE_____EUR___ _ News.txt
23. ORANGE____EUR___ _ News.txt
24. PAYPOINT ORD 1_3P____GBX___ _ News.txt
25. PVA TEPLA AG O_N_____EUR___ _ News.txt
26. RELIEF THERAPEUTICS HLDG AG RLFTF _ News.txt
27. RENAULT____EUR___ _ News.txt
28. SAI_TECH GLOBAL____USD___ _ News.txt
29. SANDVIK AB _ News.txt
30. SANOFI____EUR___ _ News.txt
31. SOCIETE GENERALE____EUR___ _ News.txt
32. SPAR NORD BANK A_S _CBOE_____DKK___ _ News.txt
33. STATOIL ASA NK 2_50 _ News.txt
34. SYENSQO____EUR___ _ News.txt
35. VOLVO A FRIA _ News.txt
```

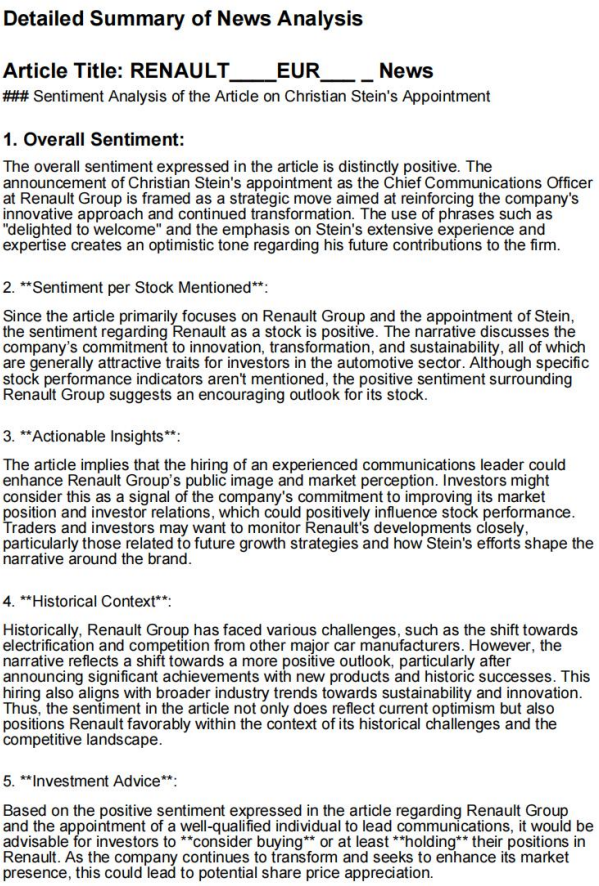
Figure 5.4.12 View Saved Articles (PyCharm Terminal)

```
34. SYENSQO____EUR___ _ News.txt
35. VOLVO A FRIA _ News.txt
Enter the number of the article you want to analyze (or 'q' to quit): 4
Analyzing the selected article...
Analysis saved to PDF: C:/Users/User/PycharmProjects/FYP1/Main/PDFs\BOSSARD N____CHF___ _ News.pdf
```

Figure 5.4.13 Analyze Saved Articles (PyCharm Terminal)

```
Choose an option: 4
Exiting the news menu.
```

Figure 5.4.14 Exit (PyCharm Terminal)



Detailed Summary of News Analysis

Article Title: RENAULT____EUR___ _ News

Sentiment Analysis of the Article on Christian Stein's Appointment

1. Overall Sentiment:

The overall sentiment expressed in the article is distinctly positive. The announcement of Christian Stein's appointment as the Chief Communications Officer at Renault Group is framed as a strategic move aimed at reinforcing the company's innovative approach and continued transformation. The use of phrases such as "delighted to welcome" and the emphasis on Stein's extensive experience and expertise creates an optimistic tone regarding his future contributions to the firm.

2. **Sentiment per Stock Mentioned**:

Since the article primarily focuses on Renault Group and the appointment of Stein, the sentiment regarding Renault as a stock is positive. The narrative discusses the company's commitment to innovation, transformation, and sustainability, all of which are generally attractive traits for investors in the automotive sector. Although specific stock performance indicators aren't mentioned, the positive sentiment surrounding Renault Group suggests an encouraging outlook for its stock.

3. **Actionable Insights**:

The article implies that the hiring of an experienced communications leader could enhance Renault Group's public image and market perception. Investors might consider this as a signal of the company's commitment to improving its market position and investor relations, which could positively influence stock performance. Traders and investors may want to monitor Renault's developments closely, particularly those related to future growth strategies and how Stein's efforts shape the narrative around the brand.

4. **Historical Context**:

Historically, Renault Group has faced various challenges, such as the shift towards electrification and competition from other major car manufacturers. However, the narrative reflects a shift towards a more positive outlook, particularly after announcing significant achievements with new products and historic successes. This hiring also aligns with broader industry trends towards sustainability and innovation. Thus, the sentiment in the article not only does reflect current optimism but also positions Renault favorably within the context of its historical challenges and the competitive landscape.

5. **Investment Advice**:

Based on the positive sentiment expressed in the article regarding Renault Group and the appointment of a well-qualified individual to lead communications, it would be advisable for investors to **consider buying** or at least **holding** their positions in Renault. As the company continues to transform and seeks to enhance its market presence, this could lead to potential share price appreciation.

Figure 5.4.15 Example output to generated PDF for New Sentiment Analysis

Each news-generated PDF presents a comprehensive analysis of articles related to specific companies. It includes sentiment analysis, actionable insights, and investment

recommendations based on recent news, providing users with detailed assessments of company performance and future outlook

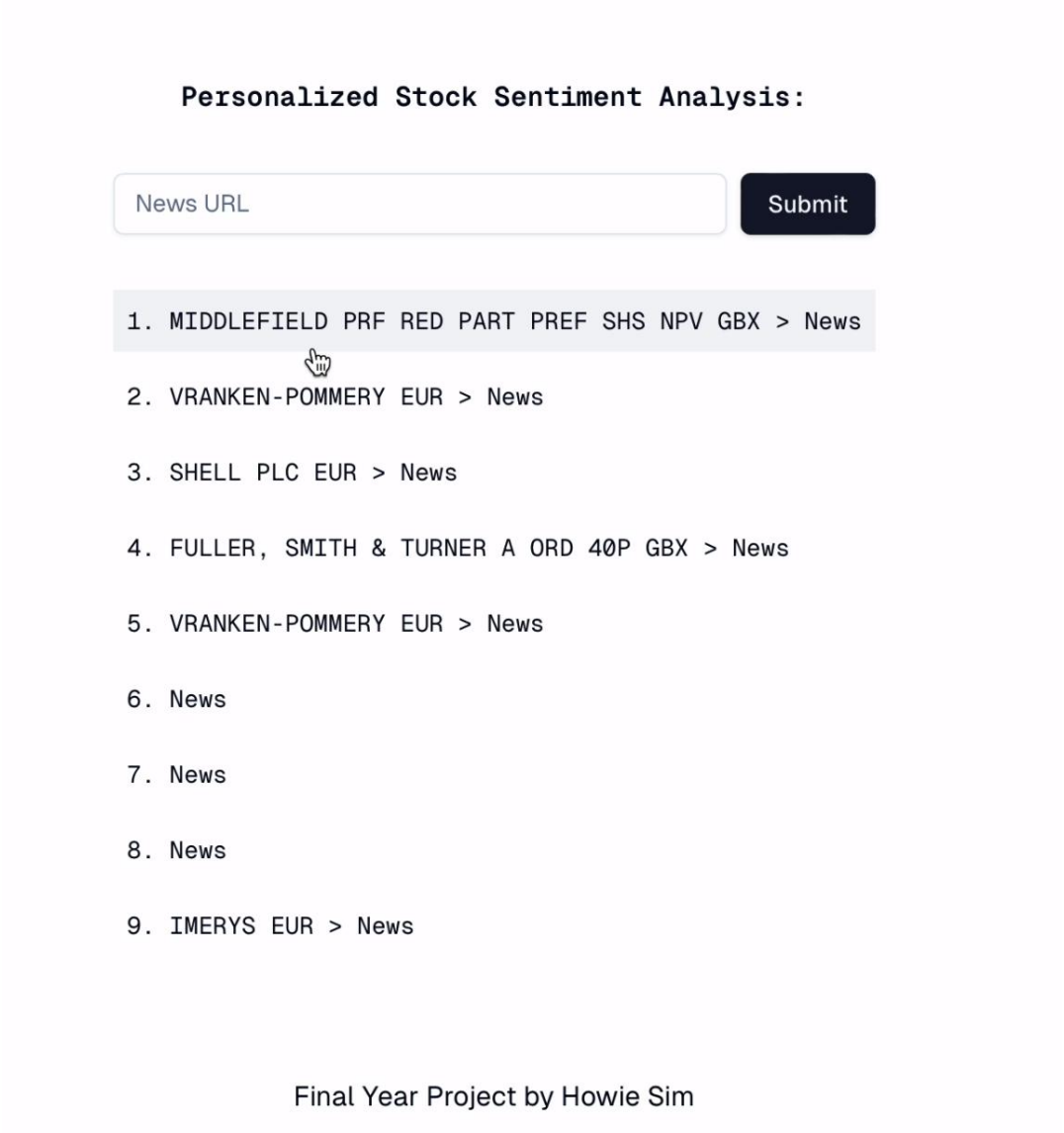


Figure 5.4.15 Search Bar Interface for News (Webpage)

5.5 Implementation Issues and Challenges

The development and implementation of the system faced several issues and challenges, which needed to be addressed to ensure a functional and efficient workflow. One of the primary challenges was optimizing the integration of the GPT-4 API for different analysis tasks. Balancing the cost of API usage with the depth of analysis required careful planning, especially when handling large datasets from multiple YouTube channels and news articles. The system had to be configured to allow users to switch between models like GPT-4-mini and GPT-4-turbo based on the complexity of the data, while maintaining a smooth user experience and minimizing processing time.

Another significant issue was the handling of real-time data scraping from YouTube and news websites. Frequent changes in website structures, especially for news sources, led to scraping failures or incomplete data extraction. This required ongoing adjustments to the scraping algorithms to ensure robustness and accuracy in gathering the required content. Additionally, ensuring data consistency across different sources was challenging, especially when comparing sentiment analysis results from YouTube and news platforms. Inconsistent data formats and varying levels of content depth required preprocessing efforts to standardize the data before analysis.

There were also challenges related to generating and merging PDF reports. The process of combining multiple PDF files for the comparative sentiment analysis often led to formatting issues, especially when handling different types of analysis outputs. Ensuring proper alignment, text wrapping, and logo placement in the final merged PDF report required significant debugging and customization. Lastly, The lack of a user interface in the current system is due to the unexpected loss of key FYP files, which required the project to be rebuilt from the ground up. This setback significantly impacted the development timeline, and the rebuilding of core components took precedence over the creation of a full-fledged interface. Although the system's primary functionalities—such as custom watchlists, and analysis options—were maintained, focus was placed on ensuring these features operated reliably within a simplified structure.

CHAPTER 6

System Evaluation and Discussion

6.1 System Testing and Performance Metrics

The reliability of the system, ensuring that the analysis is fact-based rather than speculative, was rigorously tested. Each analysis required approximately USD 0.10 to 0.12, and evaluations were conducted using three renowned YouTube stock content creators available in the system's predefined channels. Each creator was analyzed from three distinct perspectives: optimistic, pessimistic, and skeptical. Tests were conducted at two-week and one-week intervals for YouTube content to monitor consistency over time. This approach resulted in a total of 10 individual tests, providing a comprehensive assessment of the system's accuracy and consistency across various scenarios.

Additionally, the system was extended to include sentiment analysis for news articles, which was verified with five different news sources. The analysis of news content was performed over a one-week interval. The inclusion of news introduced challenges due to the varying tone and complexity of the articles. Adjustments were made to the system's pre-processing to ensure that both video and text content could be analyzed effectively, while maintaining consistency in sentiment interpretation across both formats.

6.2 Testing Setup and Result

Testing Result Tables for Short-term Prospect.
Green Color meaning the analysis was correct otherwise will be Red Color.

YouTube Creator - Parkev Tatevosian

Company (Stock)	System's Analysis (Short-term)	Actual Movement From Yahoo Finance (2 weeks)

SMCI (Negative)	Discussion focused on soaring revenue amid falling profit margins, a high dependency on a single large customer, and negative cash flow from operations. The company's stock has been volatile and adversely affected post-earnings due to these negative factors.	- $456.69 -> $427.44 (-6.40%)
NVDA (Positive)	Contrasted with SMCI, Nvidia is noted for robust sales at full pricing without offering discounts, bolstering its stock performance. This serves as a benchmark for quality revenue versus SMCI's current situation.	- $120.99 -> $116.85 (-3.42%)
PLTR (Optimistic)	The sentiment towards Palantir stock is overwhelmingly positive. The mention of "government revenue" and "revenue growth increased by more than 20%" highlights strong financial performance,	- $31.58 -> $34.82 (+10.26%)

	which is likely to attract investor interest. Additionally, the reference to the improvement in "net revenue retention rate" indicates a robust business model and customer loyalty, which further emphasizes positivity regarding Palantir's future prospects.	

Table 6.2.1 2 weeks verification - Parkev Tatevosian

YouTube Creator - Everything Money

Company (Stock)	System's Analysis (Short-term)	Actual Movement From Yahoo Finance (1 weeks)
OXY (Negative - Cautious Skepticism)	- The speaker acknowledges Buffett's significant holdings and purchases in OXY but expresses confusion over low return on invested capital and volatility, noting that this stock doesn't appeal to him personally. The positive note of a decent	- $56.36 -> $52.49 (-6.87%)

	dividend yield and free cash flow does exist, yet overshadowed by concerns about capital returns. - Occidental Petroleum, HEI Aviation Corp, and Sirius XM should be approached with caution given the highlighted uncertainties and challenges.	
HEI Avitation Corp (Negative Bias)	- The speaker is perplexed about Buffett's reasons for investing, citing high debt levels, low ROI, and volatility. There seems to be a lack of understanding regarding the stock's attractiveness, suggesting a negative bias. - Caution should be exercised regarding Occidental Petroleum, HEI Aviation, and Sirius XM until clearer visibility into their financial health and	- $255.43 -> $256.23 (+0.31%) - Was decreasing in share price until a sudden rebound of +2.8%.

	strategic direction emerges.	
ULTA (Optimistic)	- There's a shift in the tone here, as the speaker expresses understanding and approval of Ulta's business model. Factors like strong same-store sales growth and strategic positioning in retail contribute positively, suggesting a potential buy or hold for investors. - Ulta Beauty and Chubb Insurance appear to be the most favorable investments based on the speaker's analysis. Investors may consider these for potential growth due to their strong fundamentals and market positions.	- $352.95 -> $379.76 (+ 7.60%)
SIRI (Neutral to slight Negative)	- The speaker questions the logic behind Berkshire's stock movement and the	- $32.57 - > $27.21 (-19.70%)

	sustainable growth of Sirius. While acknowledging the business, there's confusion about share sell-offs, hinting at a lack of confidence. - Caution should be exercised regarding Occidental Petroleum, HEI Aviation, and Sirius XM until clearer visibility into their financial health and strategic direction emerges.	
CB (Positive)	- The speaker has a favorable view of Chubb, referencing Berkshire's expertise in insurance and long-term viability in the sector. This stock stands out as a strong buy recommendation. - Ulta Beauty and Chubb Insurance appear to be the most favorable investments based on the	- $284.0 -> $291.69 (+2.71%)

	speaker's analysis. Investors may consider these for potential growth due to their strong fundamentals and market positions.	

Table 6.2.2 1 week verification - Everything Money

YouTube Creator - Stock Moe

Company (Stock)	System's Analysis (Short-term)	Actual Movement From Yahoo Finance (1 weeks)
BAC (Negative)	- The sentiment is largely negative due to Warren Buffett's continued selling. The speaker views this as a significant red flag and expresses concern that Buffett may continue reducing his stake to below 10%. References to substantial selling pressure indicate a bearish outlook tied to Buffett's actions, suggesting that the price might decline further. - Sell or hold off on buying until clearer indicators suggest stabilizing prices or renewed investment by key figures like Buffett.	- $40.70 -> $39.58 (-2.75%)
BRK (Positive)	- Sentiment around Berkshire Hathaway	- $478.25 -> $465.04 (-2.76%) - Was decreasing in share price until a

	is much more positive. Despite concerns about Buffett's selling, there's recognition that the company is well-managed, outperforming the S&P; 500, and generally regarded as a strong investment. The speaker mentions its solid PE ratio, suggesting it is fairly valued, which contributes to a more bullish sentiment. - Buy or hold as a long-term investment given its historical performance, sound management, and current evaluations being relatively strong compared to the broader market.	sudden rebound of +2.8%.

Table 6.2.3 1 week verification - Stock Moe

News

Company (Stock)	System's Analysis (Short-term)	Actual Movement From Yahoo Finance (1 weeks)
TSLA (Negative for Short Term)	1. Tesla's Production: - Sentiment: Mixed - Details: While there is optimism about future production capacities, ongoing delays have raised concerns among investors about short-term performance. 2. Regulatory Scrutiny: - Sentiment: Negative - Details: Increased regulatory scrutiny has resulted in fines and stricter compliance measures, which could impact Tesla's profitability. **<u>Actionable Insights:</u>** - For Investors: Consider the long-term potential of Tesla's innovation, but be aware of short-term risks associated with	- $212.00 -> $216.08 (+1.92%)

	production challenges and regulatory issues.	
Latin Metals (Positive)	- The overall sentiment of the article is **positive**, with a strong emphasis on the potential for growth and development regarding Latin Metals Inc. The announcement of acquiring additional land, the continuing exploration activities, and the specificity of mineral findings all contribute to an optimistic outlook on the company's future prospects. The positive sentiment is further bolstered by the proactive steps the company is taking regarding funding and resource management. **<u>Actionable Insights:</u>** - Monitoring Future Developments: Investors	- $0.0692 -> $0.0602 (-13.00%)

	should keep an eye on the ongoing exploration results at the Auquis Project, particularly the outcomes from additional sampling at the Rose zone and the Blanco skarn target. - Assessing Financial Health: The loans, although related-party transactions, indicate management's commitment to short-term stability while pursuing long-term goals. - Consideration for Joint Ventures: The company's Prospect Generator model may offer opportunities for investors to capitalize on discoveries without infusing direct capital into drill-based exploration.	

RLFTF (Positive)	- The overall sentiment expressed in the article regarding Relief Therapeutics is **positive**. The announcement of the interim results from the PKU GOLIKE study highlights promising outcomes that suggest potential advantages of the treatment compared to existing options. The anticipation surrounding the upcoming presentation at a significant medical symposium indicates an optimistic outlook for both the resolution of the study results and the commercial prospects for the company's products. **<u>Actionable Insights:</u>** - Invest in Relief Therapeutics: Given the positive interim data and potential market applications, investors might consider taking a	- $1.17-> $2.28 (+94.87%)

	position in Relief Therapeutics. - Monitor Further Developments: Continuous monitoring of subsequent announcements regarding final study results will be critical for assessing long-term investment viability.	
MRX.DE (Positive)	- The overall sentiment of the article is decidedly **positive**. The announcement focuses on Mister Spex's strategic expansion into the premium segment with the addition of the Jacquemus sunglasses collection. Key phrases like "modern design and high quality," "significant step," and "highest standards of design and quality" reinforce this positivity. The sentiment is strong, reflecting	- $2.57 -> $2.43 (-5.45%)

	optimism about the company's growth, profitability goals, and brand positioning. **<u>Actionable Insights:</u>** - The article indicates that investors and stakeholders should consider Mister Spex's growth strategy and adaptability in the fashion and eyewear industry. This may be a signal for investors to monitor the company's performance closely and evaluate the potential for stock value appreciation, especially as it enhances its product offerings.	
Renault	The overall sentiment expressed in the article is distinctly positive. The announcement of Christian Stein's appointment as the Chief Communications Officer at Renault Group is	- $42.57 -> $38.49 (-9.58%)

	framed as a strategic move aimed at reinforcing the company's innovative approach and continued transformation. The use of phrases such as "delighted to welcome" and the emphasis on Stein's extensive experience and expertise creates an optimistic tone regarding his future contributions to the firm. **<u>Actionable Insights:</u>** The article implies that the hiring of an experienced communications leader could enhance Renault Group's public image and market perception. Investors might consider this as a signal of the company's commitment to improving its market position and investor relations, which could positively influence stock	

	performance. Traders and investors may want to monitor Renault's developments closely, particularly those related to future growth strategies and how Stein's efforts shape the narrative around the brand.	

Table 6.2.4 1 week verification of News

All the information generated by the system is based on the inputs provided during the 10 testing scenarios. The YouTube analysis yielded a 70% success rate, with 7 out of 10 predictions being correct. It only demonstrated a incorrect prediction of 3 out of 10 , suggesting that the system performs reliably in processing video content and giving profitable actionable insights to investors.

On the other hand, the sentiment analysis for news articles presented more significant challenges, with only 1 out of 5 predictions being accurate. The lower accuracy for news can be attributed to several factors: the inherent variability in tone and structure of news articles, the presence of nuanced language, and the complexity of news contexts, which can differ significantly from the more straightforward commentary in YouTube videos. Additionally, news articles often cover a broader scope of events, policies, and economic factors, making it harder for the system to extract clear sentiment aligned with stock performance predictions. Despite this, the system's ability to handle such diverse input formats demonstrates its flexibility, and with further refinements, improvements in news analysis are expected.

6.3 Objectives Evaluation

This project has successfully developed a new, efficient method for making informed stock investments, offering an innovative alternative to traditional approaches like self-learning or hiring fund managers. The YouTube analysis, with a 70% accuracy rate, highlights the system's capability to provide reliable insights based on video content. The more complex nature of news articles, reflected in the lower accuracy rate, points to opportunities for enhancement. Justifying the discrepancy, it is clear that news articles often employ more nuanced and layered language, requiring further optimization of the sentiment analysis model.

By simplifying the complex process of self-learning—usually a time-consuming endeavor involving market research, financial literature, and educational courses—the system effectively streamlines this process. Moreover, the dual purpose of the system as both an analytical tool and an educational resource makes it an accessible and valuable asset for beginners. This practical solution reduces the financial and time investments traditionally required to enter the stock market.

Additionally, our system offers an alternative to hiring fund managers, whose fees can be prohibitively high. Through a pay-as-you-go model, users are empowered to conduct detailed analyses of companies at a fraction of the cost, paying only for GPT API usage. This feature, along with its long-term investment recommendations backed by robust analysis, positions the system as a valuable resource for both novice and experienced investors.

Chapter 7

Conclusion and Recommendation

7.1 Conclusion

Investing in the stock market can be a rewarding strategy for increasing wealth, but it often demands extensive research and analysis of various companies. Traditional stock forecasting tools primarily rely on quantitative data, which, while useful, can fall short by not fully capturing the impact of qualitative factors such as market sentiment, economic shifts, and political events. This limitation arises because the stock market is influenced by more than just numerical data; it is also shaped by unpredictable variables.

This project aims to address these limitations by leveraging sentiment analysis to forecast stock prices using qualitative data. By focusing on user-selected YouTube content related to stock analysis and recent company news articles, this project explores a novel approach to stock price prediction. Users will have the opportunity to curate their own watchlists of preferred stock analysis YouTubers, allowing for a personalized sentiment analysis experience. Additionally, sentiment analysis will be applied to news articles from the past two weeks to provide a current perspective on company performance.

The core of this project is to generate comparative reports that analyze the sentiment from both YouTube content and news articles, and evaluate their correlation with stock price movements and historical trends. By comparing sentiment-based predictions with historical stock data, this project aims to assess the reliability and potential of qualitative data in forecasting stock prices. The insights gained from this comprehensive analysis will illuminate both the strengths and limitations of sentiment analysis, offering valuable recommendations for future research and application in the field of stock market predictions.

7.2 Recommendation

Due to the time constraints, there are a few areas I wish future candidate can explore on; if there's any to continue this project.

1. **Implement Real-Time Data Integration**

Integrate APIs that provide real-time financial news, stock prices, and social media sentiment. Use data streaming technologies such as Apache Kafka or AWS Kinesis to handle live data feeds efficiently. Ensure that the system can process and analyze this data in near real-time to maintain the relevance and timeliness of predictions.

2. **Employ Advanced Statistical Analysis**

Integrate advanced statistical techniques to better understand the correlation between sentiment and stock price movements. Use econometric models such as Vector Autoregressive Models (VAR) or Granger Causality Tests to explore causal relationships. Incorporate time-series analysis tools like ARIMA (AutoRegressive Integrated Moving Average) to model and forecast stock price trends.

3. **Enhance Data Preprocessing and Normalization**

Develop advanced data preprocessing pipelines to handle diverse data sources, including text normalization, entity recognition, and sentiment tagging. Implement text cleaning procedures to remove noise and irrelevant information from YouTube transcripts and news articles. Use tools like SpaCy for entity recognition and NLTK for text processing to improve the quality of input data.

4. **Enhance Data Visualization**

Implement advanced data visualization tools to present comparative sentiment analysis results effectively. Use libraries like D3.js or Plotly to create interactive and dynamic visualizations. Develop dashboards using tools such as Tableau or Power BI to provide users with comprehensive and intuitive views of sentiment data and stock price correlations.

REFERENCES

[1] S. K. Khatri and A. Srivastava, “Using sentimental analysis in prediction of Stock Market Investment,” 2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2016. doi:10.1109/icrito.2016.7785019

[2] U. Gupta, “GPT-InvestAR: Enhancing stock investment strategies through annual report analysis with large language models,” arXiv.org, https://arxiv.org/abs/2309.03079

[3] , “Enhancing sentiment analysis based investment by large language models in Japanese stock market,” SSRN Electronic Journal, 2023. doi:10.2139/ssrn.4511658

[4] “Which Artificial Intelligence Algorithm Better Predicts the Chinese Stock Market?,” ieeexplore.ieee.org. https://ieeexplore.ieee.org/document/8419702

[5] A. Zheng and J. Jin, “Using AI to Make Predictions on Stock Market,” 2017. Available: https://cs229.stanford.edu/proj2017/final-reports/5212256.pdf

[6] R. Chopra and G. D. Sharma, “Application of Artificial Intelligence in Stock Market Forecasting: A Critique, Review, and Research Agenda,” Journal of Risk and Financial Management, vol. 14, no. 11, p. 526, Nov. 2021, doi: https://doi.org/10.3390/jrfm14110526

[7] F. G. D. C. Ferreira, A. H. Gandomi, and R. T. N. Cardoso, “Artificial Intelligence Applied to Stock Market Trading: A Review,” IEEE Access, vol. 9, no. 2169–3536, pp. 30898–30917, Feb. 2021, doi: https://doi.org/10.1109/access.2021.3058133.

[8] C. Zhang et al., “When AI Meets Finance (StockAgent): Large Language Model-based Stock Trading in Simulated Real-world Environments,” arXiv.org, 2024. https://arxiv.org/abs/2407.18957 (accessed Aug. 22, 2024).

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Trimester 3, Year 3	**Study week no.: 2**
Student Name & ID: Sim Kah Hoe (21ACB06036)	
Supervisor: Prof Ts Dr. Liew Soung Yue	
Project Title: Accelerated Personalized Stock Sentiment Analysis: Leveraging LLMs for YouTuber Content and News Articles	

1. WORK DONE

None

2. WORK TO BE DONE

Explore and research new needed and good features that can be possibly added.

3. PROBLEMS ENCOUNTERED

None

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: Sim Kah Hoe (21ACB06036)	
Supervisor: Prof Ts Dr. Liew Soung Yue	
Project Title: Accelerated Personalized Stock Sentiment Analysis: Leveraging LLMs for YouTuber Content and News Articles	

1. WORK DONE

Explored and researched new needed and good features that can be possibly added.

2. WORK TO BE DONE

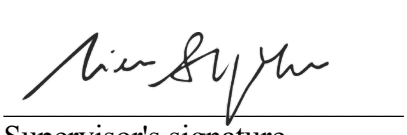
Start building each functions.

3. PROBLEMS ENCOUNTERED

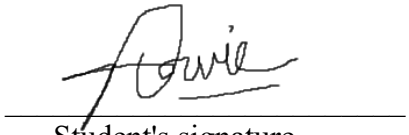
None

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: Sim Kah Hoe (21ACB06036)	
Supervisor: Prof Ts Dr. Liew Soung Yue	
Project Title: Accelerated Personalized Stock Sentiment Analysis: Leveraging LLMs for YouTuber Content and News Articles	

1. WORK DONE

Start building each functions.

2. WORK TO BE DONE

Verification & Testing relevancy on the new functions added

3. PROBLEMS ENCOUNTERED

Laptop malfunctioning, frequent BSOD/crashing/freezing.

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: Sim Kah Hoe (21ACB06036)	
Supervisor: Prof Ts Dr. Liew Soung Yue	
Project Title: Accelerated Personalized Stock Sentiment Analysis: Leveraging LLMs for YouTuber Content and News Articles	

1. WORK DONE

Verification & Testing relevancy on the new functions added.
Send laptop to fix.

2. WORK TO BE DONE

Verification on stock prediction accuracy & Testing old and new functions.

3. PROBLEMS ENCOUNTERED

Laptop malfunctioning, frequent BSOD/crashing/freezing.

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: Sim Kah Hoe (21ACB06036)	
Supervisor: Prof Ts Dr. Liew Soung Yue	
Project Title: Accelerated Personalized Stock Sentiment Analysis: Leveraging LLMs for YouTuber Content and News Articles	

1. WORK DONE

Verification on stock prediction accuracy & Testing old and new functions.

2. WORK TO BE DONE

Streamline the functions added in FYP2.

3. PROBLEMS ENCOUNTERED

Laptop fix but FYP file corrupted. Need to rebuild the system.

4. SELF EVALUATION OF THE PROGRESS

Not the best but not the worst, still can handle.

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: Sim Kah Hoe (21ACB06036)	
Supervisor: Prof Ts Dr. Liew Soung Yue	
Project Title: Accelerated Personalized Stock Sentiment Analysis: Leveraging LLMs for YouTuber Content and News Articles	

1. WORK DONE

Rebuilt the system.

2. WORK TO BE DONE

Report and front end (If possible)

3. PROBLEMS ENCOUNTERED

None

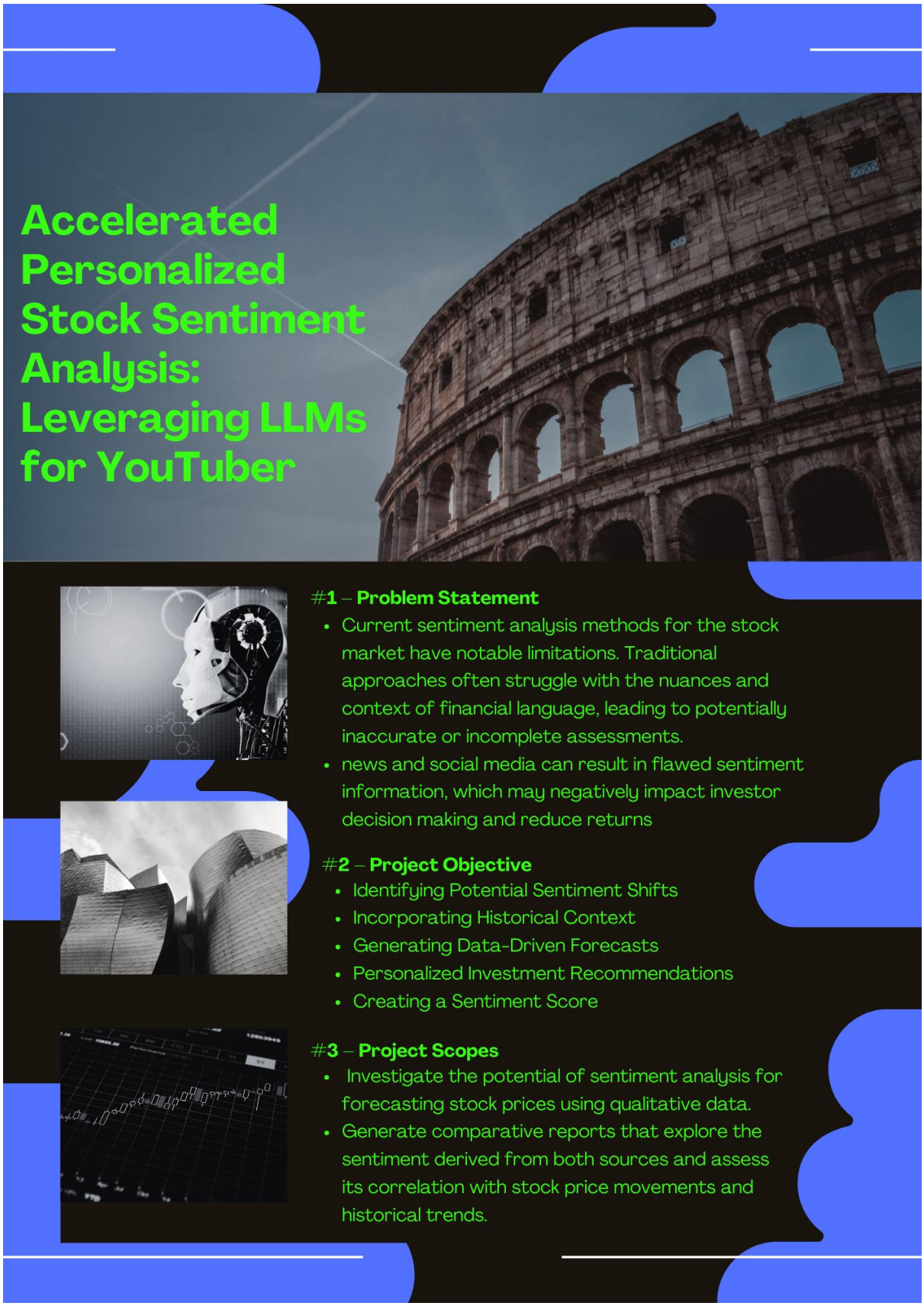
4. SELF EVALUATION OF THE PROGRESS

Good.

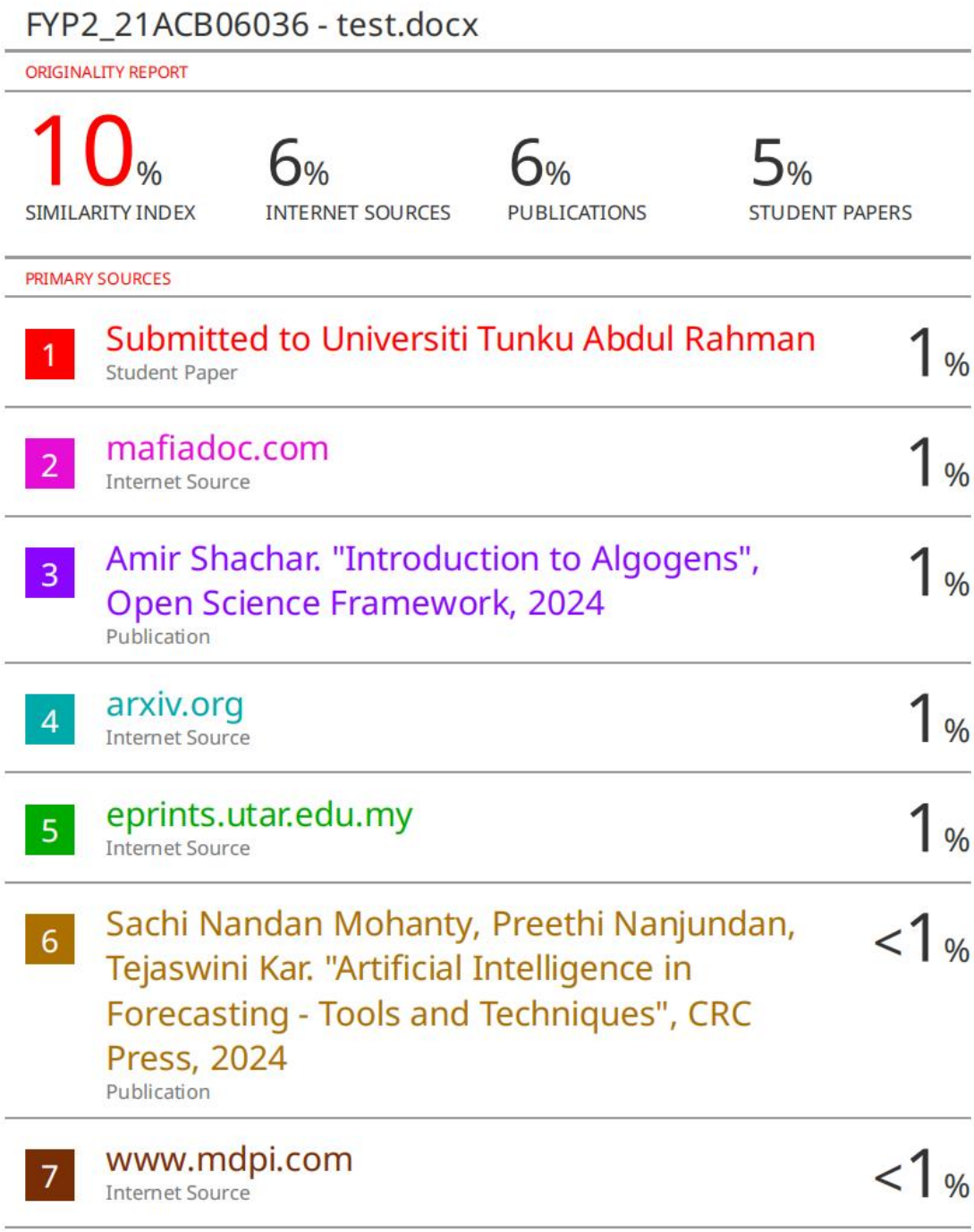
Supervisor's signature

Student's signature

POSTER


Accelerated Personalized Stock Sentiment Analysis: Leveraging LLMs for YouTuber
#1 – Problem Statement
• Current sentiment analysis methods for the stock market have notable limitations. Traditional approaches often struggle with the nuances and context of financial language, leading to potentially inaccurate or incomplete assessments.
• news and social media can result in flawed sentiment information, which may negatively impact investor decision making and reduce returns
#2 – Project Objective
• Identifying Potential Sentiment Shifts
• Incorporating Historical Context
• Generating Data-Driven Forecasts
• Personalized Investment Recommendations
• Creating a Sentiment Score
#3 – Project Scopes
• Investigate the potential of sentiment analysis for forecasting stock prices using qualitative data.
• Generate comparative reports that explore the sentiment derived from both sources and assess its correlation with stock price movements and historical trends.

PLAGIARISM CHECK RESULT



FYP2_21ACB06036 - test.docx

ORIGINALITY REPORT

10%	6%	6%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universiti Tunku Abdul Rahman Student Paper	1%
2	mafiadoc.com Internet Source	1%
3	Amir Shachar. "Introduction to Algogens", Open Science Framework, 2024 Publication	1%
4	arxiv.org Internet Source	1%
5	eprints.utar.edu.my Internet Source	1%
6	Sachi Nandan Mohanty, Preethi Nanjundan, Tejaswini Kar. "Artificial Intelligence in Forecasting - Tools and Techniques", CRC Press, 2024 Publication	<1%
7	www.mdpi.com Internet Source	<1%

8	Rafid Ullah, Hishamuddin Bin Ismail, Mohammad Tariqul Islam Khan, Ali Zeb. "Nexus between Chat GPT usage dimensions and investment decisions making in Pakistan: Moderating role of financial literacy", Technology in Society, 2024 Publication	<1 %
9	www.fastercapital.com Internet Source	<1 %
10	papers.ssrn.com Internet Source	<1 %
11	tudr.org Internet Source	<1 %
12	"Advances in Computational Intelligence Systems", Springer Science and Business Media LLC, 2024 Publication	<1 %
13	Submitted to Swinburne University of Technology Student Paper	<1 %
14	Nabanita Das, Bikash Sadhukhan, Rajdeep Chatterjee, Satyajit Chakrabarti. "Integrating sentiment analysis with graph neural networks for enhanced stock prediction: A comprehensive survey", Decision Analytics Journal, 2024 Publication	<1 %

15	Biswadip Basu Mallik, Gunjan Mukherjee, Rahul Kar, Aryan Chaudhary. "Deep Learning Concepts in Operations Research", Routledge, 2024 Publication	<1 %
16	Joel Jebadurai Devapitchai, Krishnapriya S. V., Karuppiah S. P., William Robert P., S. Saranya. "chapter 6 Using AI-Driven Decision-Making Tools in Corporate Investment Planning", IGI Global, 2024 Publication	<1 %
17	Submitted to Symbiosis International University Student Paper	<1 %
18	www.livemint.com Internet Source	<1 %
19	Submitted to University of Liverpool Student Paper	<1 %
20	Submitted to City University Student Paper	<1 %
21	Claire E. Robertson, Nicolas Pröllochs, Kaoru Schwarzenegger, Philip Pärnamets, Jay J. Van Bavel, Stefan Feuerriegel. "Negativity drives online news consumption", Nature Human Behaviour, 2023 Publication	<1 %
	Submitted to University of Bolton	

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)	Sim Kah Hoe
ID Number(s)	21ACB06036
Programme / Course	BACHELOR OF COMPUTER SCIENCE (HONOURS)
Title of Final Year Project	Accelerated Personalized Stock Sentiment Analysis: Leveraging LLMs for YouTuber Content and News Articles

Similarity	**Supervisor's Comments (Compulsory if parameters of originality exceeds the limits approved by UTAR)**
Overall similarity index: **__10_____ % Similarity by source** Internet Sources: _____6_____% Publications: _____6_____ % Student Papers: _____5_____ %	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: 13/9/2024

Signature of Co-Supervisor

Name: _____

Date: _____

UNIVERSITI TUNKU ABDUL RAHMAN

FACULTY OF INFORMATION & COMMUNICATION TECHNOLOGY
(KAMPAR CAMPUS)

CHECKLIST FOR FYP2 THESIS SUBMISSION

Student Id	21ACB06036
Student Name	Sim Kah Hoe
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: 12 September 2024