



GRADUATE CERTIFICATE
ISY 5005
INTELLIGENT SOFTWARE AGENTS (ISA)
PROJECT MODULE
PROJECT REPORT

INVESTOR INTELLIGENCE AGENT

Yatharth Mahesh Sant (A0286001R)
Kristofer Roos (A0285949A)
Goh Min Hua (A0285810A)
Zhang Eunice (A0287088L)
Chua Kian Yong Kenny (A0056377W)

Table of Contents

1.	EXECUTIVE SUMMARY	5
2.	BUSINESS CASE AND MARKET RESEARCH.....	7
2.1	Market Needs and Opportunities	7
2.2	Industry Trends	7
2.3	Quantitative and Qualitative Impacts	8
2.4	Competitive Landscape.....	9
2.4.1	Competitors	9
2.4.2	SWOT Analysis.....	10
	Strengths:	10
	Weaknesses:	10
	Opportunities:.....	11
	Threats:.....	11
2.5	Value Proposition, Competitive Positioning, and Benchmarking	11
2.5.1	Value Proposition	11
	Deep Learning and AI Integration for Personalized Analytics	11
	Robotic Process Automation and Real-Time Data for Timely Decision-Making	12
	Comprehensive ESG and Regulatory Monitoring for Informed Investments	12
	User-Centric Design and Accessibility Enhances User Experience.....	12
	Customizable and Scalable Solutions Meet Diverse Needs	12
2.5.2	Competitive Positioning	12
2.5.3	Benchmarking	13
2.6	Pricing Strategy.....	14
2.6.1	Investors and Consulting Firms	14
2.6.2	Corporate Entities.....	14
3.	SYSTEM DESIGN AND MODEL	16
3.1	Architecture Overview	16
3.2	System Overview / Detailed Workflow	17
3.3	Technical Stack	18

3.3.1	Frontend:	18
3.3.2	Backend:	18
3.3.3	Database:	18
3.4	Data Management	18
3.4.1	Data Sources.....	18
3.4.2	Data Preparation	18
4.	SYSTEM DEVELOPMENT AND IMPLEMENTATION	21
4.1	Tools / Techniques	21
4.1.1	Gemma-2B Supervised Fine Tuning	21
	Main Components:.....	23
4.1.1	Assessing Key Company Information and Financial Metrics	23
4.1.2	ESG Integration in Financial Analytics	26
4.1.3	Real-Time News Informer Module	28
	News Scraper & Analysis Module Overview	28
	Integration of NewsAPI for Comprehensive Data Collection.....	28
	Robotic Process Automation for Dynamic News Scraping.....	28
	Enhancing Data Integrity Through Duplicate Removal	29
	Segmented Sentiment Analysis	29
	Sentiment Summarization and Interpretation	29
4.1.4	Neuro Fuzzy Model.....	30
4.1.5	Backend	31
4.1.6	Frontend and Others	31
4.2	Challenges and Solutions	35
4.2.1	Nature of Streamlit framework	35
4.2.2	Parsing User Input for the NLP Chatbot	35
5.	FINDINGS AND DISCUSSION.....	36
5.1	Risk Analysis	36
5.1.1	Technical Risks	36
5.1.2	Compliance Risks.....	36
5.1.3	Market and Operational Risks.....	37
5.1.4	Financial and Reputational Risks	38

5.1.5	ESG Analysis Risks	38
5.2	Roadmap for Future	38
5.2.1	Future Considerations.....	38
	Web Scraping via TagUI	39
	ESG Analysis	39
	Enhancing Competitor Analysis	39
	Developing Benchmarks for Evaluating Companies.....	40
	Integrated Predictive Modelling	40
	Dynamic Scenario Analysis Tool	40
	Real-time Regulatory Change Tracker	41
6.	CONCLUSIONS.....	42
	APPENDIX A – INSTALLATION AND USER GUIDE	43
	Installation.....	43
	Directory Layout.....	43
	Brief Description.....	43
	Instructions on how to run the Code.....	43
	System Requirements	44
	Additional Resources	44
	User Guide	45
	REFERENCES.....	48

1. EXECUTIVE SUMMARY

The Investor Intelligence Agent, developed for the ISY 5005 Intelligent Software Agents course, represents a transformative step in financial advisory services. This platform provides the aggregation of several complementary and novel data sources, generating timely insights which are currently only possible with expensive investment analyst resources. These automated scraping and analysis modules provide a new frontier for fast, cheap, accurate, and customized investment insights at a large scale.

At the core of the Investor Intelligence Agent are several innovative features designed to optimize financial analysis and investment strategies. The platform integrates Gemma-2B for supervised fine-tuning and ANFIS (Adaptive Neuro-Fuzzy Inference System), both of which provide customized risk assessments and predictive analytics. These models learn continuously from vast amounts of market data, adapting to new information to deliver precise and actionable insights. Furthermore, the platform employs TagUI for robust Robotic Process Automation (RPA), ensuring efficient and timely data scraping which is critical for maintaining up-to-the-minute financial news and company-specific data essential for informed investment decisions. Complementing these features is the comprehensive ESG integration, utilizing specially fine-tuned BERT models for environmental, social, and governance factors. This integration not only aids investors in assessing the sustainability and ethical impact of their investments but also aligns with modern regulatory and social standards, offering a multidimensional analysis that transcends traditional financial metrics. These features collectively underscore the platform's commitment to providing a high-value, insightful, and regulatory-compliant financial advisory service.

The architecture of the Investor Intelligence Agent is engineered for robustness and flexibility, incorporating a Python-based backend that handles complex data processing tasks with high efficiency. The frontend, powered by Streamlit, provides an interactive user experience, allowing for intuitive engagement with the system's features. Data integrity and security are maintained through a SQLite database, which ensures that all user data and insights are securely stored. This technical framework supports a scalable and modular design, which is crucial for real-time data processing and enabling personalized user customizations, making the platform adaptable to the evolving needs of its users.

The Investor Intelligence Agent revolutionizes access to sophisticated financial analytics, previously available only to large institutions, by offering these powerful tools to individual investors and mid-sized firms. The platform's value

proposition is centered on delivering personalized financial insights, where algorithms tailor advice to individual user profiles, ensuring relevance and scalability. It enhances real-time decision-making capabilities through the latest advancements in machine learning and data processing, enabling users to respond swiftly to market changes. Additionally, the platform ensures adherence to global financial regulations and offers comprehensive ESG reporting, aligning with the modern investor's focus on ethical investment practices.

Looking ahead, the Investor Intelligence Agent plans to implement significant enhancements to further empower users. The introduction of a Dynamic Scenario Analysis Tool will allow investors to stress-test their investment strategies under various market conditions, providing a robust platform for risk assessment and strategic planning. Additionally, the integration of a Real-time Regulatory Change Tracker will keep users informed of legislative changes across global markets, ensuring that investment decisions remain compliant and informed by the latest regulatory developments. These enhancements are designed to fortify the platform's capabilities, ensuring it not only meets current user needs but also adapts to future market dynamics and regulatory environments.

2. BUSINESS CASE AND MARKET RESEARCH

2.1 Market Needs and Opportunities

In today's information-intensive business environment, investors and stock portfolio advisors face the challenge of staying well-informed about the companies they are interested in. The vast amount of data available, including financial metrics, industry trends, and social sentiments, necessitates a solution to compile and analyze information comprehensively.

These companies and individuals rely on timely and accurate information to make informed decisions regarding investment strategies, portfolio management, and market trends. However, traditional methods of research are time-consuming hence creating a demand for an automated system that streamlines the process of gathering, analyzing, and presenting relevant data on companies.

The potential market for the Investor Intelligence Agent is substantial, encompassing the investment sector and related advisory services. This includes investment firms, asset managers, hedge funds, private equity companies, and strategic consulting firms.

In addition to the primary user base, there is potential for expansion into other sectors, such as corporate entities seeking to enhance their understanding of market dynamics, competitor analysis, and industry trends to stay competitive in today's dynamic business environment.

2.2 Industry Trends

The financial services industry is witnessing a rapid transformation driven by several pivotal trends. Technological advancements such as hybrid cloud technology are being increasingly adopted by financial institutions to improve scalability, functionality, and regulatory compliance. This technological shift facilitates efficient management of workloads and compliance activities across decentralized platforms (IBM, 2024). Concurrently, the rise of cybersecurity threats has prompted an enhanced focus on robust cybersecurity frameworks. Financial institutions are investing heavily in sophisticated risk management tools to safeguard against these evolving threats and maintain consumer trust (IBM, 2024). Additionally, generative AI is revolutionizing the sector by introducing dynamic new product offerings and reshaping customer interactions, setting a new standard for service provision in the industry (IBM, 2024).

Sustainability initiatives are also gaining traction, with a growing emphasis on Environmental, Social, and Governance (ESG) practices. Financial institutions are increasingly committing to green initiatives, which not only comply with regulatory demands but also align with investor expectations for sustainable financial products (IBM, 2024). The regulatory landscape itself is becoming more stringent, compelling institutions to enhance their governance frameworks to remain compliant in a rapidly evolving environment (PwC, 2024). In response to consumer demand for greater control over their financial data, the expansion of open banking continues, facilitated by secure APIs that enhance financial data sharing and foster innovation (IBM, 2024).

Furthermore, the resurgence of interest in digital currencies and blockchain technology underscores a significant shift towards incorporating these technologies into mainstream financial practices. Blockchain is particularly valued for its ability to securely manage transactions and integrate with other digital currencies like Bitcoin, which is seeing increased acceptance as a legitimate form of payment (Business Partner Magazine, 2024). Lastly, digitalization in asset management is streamlining operations, enhancing performance, and facilitating proactive risk management through the use of digital tools like IoT and advanced data analytics, which enable deeper strategic insights (Business Partner Magazine, 2024).

These trends collectively indicate a shift towards more integrated, technologically advanced, and sustainability-focused financial services, necessitating that institutions adapt to stay competitive and responsive to global economic and environmental challenges.

2.3 Quantitative and Qualitative Impacts

The artificial intelligence driven financial advisory market is demonstrating substantial growth with significant implications for both market potential and business impacts. As of the latest forecasts, the Total Addressable Market (TAM) for AI-driven financial advisory tools is projected to reach a staggering \$61.3 billion by 2031. This growth is driven by a compounded annual growth rate of 22.5% from 2021, underscoring the expanding utilization of AI technologies in financial services (Allied Market Research, 2024).

Focusing on the Serviceable Available Market (SAM) within North America and Europe, these regions are identified as leaders in fintech innovation and AI adoption, contributing to approximately 50% of the global market. This estimation is predicated on the advanced technological infrastructure and high

rate of AI adoption within financial sectors across these geographies (Allied Market Research, 2024).

For our Investor Intelligence Agent, the Serviceable Obtainable Market (SOM) aims to capture 3% of the SAM over the next five years. This target aligns with the typical market penetration rates for new technological products introduced in well-established markets, reflecting a realistic and attainable growth trajectory given the current market dynamics (Allied Market Research, 2024).

The integration of AI within financial analytics is set to revolutionize cost structures and operational efficiencies. Specifically, AI is poised to reduce operational costs by approximately 25% primarily through the automation of data processing and analysis tasks. This reduction in costs is a direct result of leveraging advanced AI capabilities to streamline and enhance the accuracy of financial data management processes (McKinsey & Company, 2024).

Moreover, the decision-making process is expected to see an improvement in speed by up to 40% due to the real-time data analysis capabilities provided by AI technologies. This enhancement facilitates quicker response times to market changes and more informed decision-making, significantly benefiting financial institutions that integrate these technologies into their operational framework (McKinsey & Company, 2024).

Overall, the strategic integration of AI within the financial advisory sector holds significant promise for enhancing market responsiveness, reducing operational costs, and driving efficiency improvements, while also supporting the broader adoption of sustainable and governance-focused business practices.

2.4 Competitive Landscape

2.4.1 Competitors

In the evolving landscape of financial analytics, several major competitors dominate the market, each with distinct offerings and strategic positions. Bloomberg stands out as a leading entity in financial software, data, and media, renowned for its comprehensive analytics and extensive reporting capabilities. This positions Bloomberg at the forefront of providing in-depth market insights and data-driven solutions (Craft, 2024).

Thomson Reuters further complements the competitive environment with its specialized business information services. Known for a strong emphasis on financial data and analytics solutions, Thomson Reuters supports a wide range of business intelligence needs across various sectors, making it a pivotal player in the field (Craft, 2024).

Morningstar, another significant competitor, is acclaimed for its independent investment insights and financial data services. With a focus on integrity and reliability, Morningstar caters to investors seeking unbiased information and analysis, thereby enhancing the decision-making process for its users (Craft, 2024).

M Science leverages its expertise as a data-driven research and analytics firm, utilizing large datasets to provide detailed market insights. This approach allows M Science to deliver tailored analytical solutions that meet the specific needs of its clients (Craft, 2024).

Furthermore, comprehensive business analytics solutions are offered by giants such as Oracle Corporation, IBM Corporation, SAP SE, Microsoft Corporation, and Tibco Software Inc. These corporations play a crucial role in the broader market by delivering a variety of analytics and business intelligence tools that address complex business challenges and enhance operational efficiencies (Mordor Intelligence, 2023).

Together, these competitors form a dynamic ecosystem in the financial analytics market, ranging from dedicated financial data providers to providers of holistic business intelligence solutions. This diversity in services and expertise ensures that they cater to a broad spectrum of client needs, thus driving innovation and growth in the financial analytics sector. Given there is fierce competition and a high level of competitive players, reinforces the need for increasingly specific and targeted niche.

2.4.2 SWOT Analysis

Strengths:

The competitive landscape in the financial analytics market is notably strong due to the presence of established companies like Bloomberg and Thomson Reuters, which are renowned for their robust analytical platforms. These platforms deliver deep market insights and data accuracy, making them indispensable tools for financial professionals (Craft.co, 2024). Additionally, companies like Morningstar are recognized for their independent and reliable investment data, contributing significantly to their strong brand recognition in the industry (Craft.co, 2024).

Weaknesses:

One of the primary challenges faced by competitors in this market includes the high operational costs associated with maintaining and updating advanced analytical platforms. These costs can impact their pricing strategies, potentially

making their services less accessible to smaller entities or emerging markets (Mordor Intelligence, 2023).

Opportunities:

There is a significant opportunity for these companies to leverage emerging technologies such as AI and machine learning to enhance their data analysis capabilities. This technological advancement not only improves the efficiency and effectiveness of the analytics provided but also positions these firms to lead in innovation within the sector (Mordor Intelligence, 2023). Furthermore, expansion into emerging markets presents a lucrative growth avenue for giants like Oracle and IBM, allowing them to increase their global footprint and tap into new customer bases (Mordor Intelligence, 2023).

Threats:

The financial analytics sector is heavily regulated, requiring companies to navigate a complex regulatory landscape that can vary significantly by region and change over time. This presents a continuous risk that can affect operational stability and market strategy (Mordor Intelligence, 2023). Moreover, the rapid pace of technological change poses a threat to established players who may struggle to innovate quickly enough to keep up with new entrants and evolving market demands (Mordor Intelligence, 2023).

These insights into the SWOT analysis underscore the dynamic nature of the competitive environment in the financial analytics market, highlighting both the challenges and opportunities that shape strategic decisions in this industry.

2.5 Value Proposition, Competitive Positioning, and Benchmarking

2.5.1 Value Proposition

Deep Learning and AI Integration for Personalized Analytics

The Investor Intelligence Agent harnesses the power of advanced AI, including Gemma-2B Supervised Fine Tuning and the Adaptive Neuro-Fuzzy Inference System (ANFIS), to provide precise, personalized financial insights and risk assessments. This platform is specifically designed to adapt to the unique financial contexts of both individuals and institutions, offering customized advice that dynamically responds to evolving market conditions. Such personalized analytics empower users to make informed decisions that align with their investment strategies and risk tolerance.

Robotic Process Automation and Real-Time Data for Timely Decision-Making

Through the integration of TagUI and a sophisticated real-time news informer module, the Investor Intelligence Agent ensures seamless access to the latest financial news and critical company data. This capability is vital for enabling users to perform swift, comprehensive analyses, which are crucial in navigating the often volatile market landscapes. The platform's ability to aggregate and process real-time data not only enhances operational efficiency but also significantly improves the timeliness and accuracy of investment decisions.

Comprehensive ESG and Regulatory Monitoring for Informed Investments

Recognizing the growing importance of sustainability and compliance in investment decisions, the platform incorporates specialized BERT models fine-tuned for environmental, social, and governance factors. This advanced ESG analysis allows investors to evaluate potential investments with an added layer of insight into sustainability and regulatory compliance. By integrating these considerations, the platform provides a more holistic view of investment opportunities, aligning financial returns with ethical and regulatory standards.

User-Centric Design and Accessibility Enhances User Experience

The platform's design prioritizes user-friendliness and accessibility, employing Streamlit alongside a robust Python backend to democratize access to sophisticated analytics. This approach makes advanced data analysis tools available to non-specialists, effectively leveling the playing field for individual investors and smaller firms. By simplifying the user interface while maintaining powerful analytical capabilities, the platform engages a broader audience and enhances the overall user experience.

Customizable and Scalable Solutions Meet Diverse Needs

The Investor Intelligence Agent is designed to be highly scalable and customizable, catering to a wide range of market participants from individual investors to large institutions. Whether users require basic data visualizations or advanced predictive modeling, the platform's flexible architecture can be tailored to meet specific needs. This scalability ensures that as user requirements grow and evolve, the platform can adapt seamlessly, providing consistent, reliable, and effective support across all levels of investment activities.

2.5.2 Competitive Positioning

The Investor Intelligence Agent positions itself uniquely within the financial analytics market by strategically focusing on small to mid-sized investment firms and individual investors. This focus not only differentiates it from giants like

Bloomberg and Thomson Reuters, who primarily cater to large institutional clients, but also addresses a significant gap in the market by tailoring its offerings to the unique needs of a segment that has been historically underserved. By providing these users with advanced, AI-driven analytical tools that were once the exclusive domain of larger entities, the platform empowers them with actionable insights and data-driven decision-making capabilities. This strategic market focus is designed to cultivate a strong brand identity around accessibility, customization, and cutting-edge technology, enhancing user engagement and fostering loyalty. The Investor Intelligence Agent leverages its technological edge to offer a competitive advantage. The platform integrates the machine learning and data processing to deliver real-time analytics and comprehensive market evaluations, setting a new standard in the quality and speed of financial analysis available to its users. This not only enhances the platform's appeal but also establishes it as a leader in innovation within the niche it serves. Continuous investment in technology and a commitment to adapting to emerging market trends keep the platform at the forefront, ensuring it remains attractive and relevant to its users. This competitive positioning is underpinned by a keen understanding of the market dynamics and a clear vision of the brand as a growth partner for its clients. By consistently aligning its developments with the needs and challenges of its target users, the Investor Intelligence Agent not only stands out from its competitors but also builds a reputation as a pivotal tool for those aiming to maximize their investment potential in a highly volatile market. This strategic alignment ensures that the platform is not just a tool, but a copilot for financial transparency and market knowledge.

2.5.3 Benchmarking

In the competitive landscape of financial analytics platforms, effective benchmarking is crucial for maintaining a competitive edge. The Investor Intelligence Agent employs a robust benchmarking strategy centered on tangible performance metrics and continuous innovation tracking. To ensure the platform delivers superior value and stays ahead of market trends, we measure key performance indicators (KPIs) such as user engagement rates, customer satisfaction scores, and retention rates. These metrics are compared against industry standards, gathered from market leaders like Bloomberg and Thomson Reuters, who are known for their high customer engagement and robust analytics platforms. This comparative analysis helps identify areas where the platform can enhance its offerings or streamline operations to boost user satisfaction and engagement. Additionally, the platform's innovation and development are benchmarked against the latest advancements in AI and machine learning within the financial services sector. By keeping a pulse on technological developments

and integrating state-of-the-art features, the Investor Intelligence Agent aims to not only meet but exceed the analytical capabilities of competitors. This proactive approach involves regularly reviewing emerging technologies and adapting quickly to incorporate them, thus ensuring the platform remains relevant and powerful for user needs.

2.6 Pricing Strategy

During the launch phase, the primary goal is to establish a robust user base and demonstrate the value of the platform. To achieve this, we planned to adopt a strategic pricing model tailored to different user segments: Investors, Consulting Firms, and Corporate Entities. This multi-faceted approach ensures that we can build a sustainable business model while providing value to a diverse set of segments.

2.6.1 Investors and Consulting Firms

Initially, consumers and consulting firms will have free access to the platform. This strategy is designed to attract a substantial number of users and encourage widespread usage. By experiencing the benefits of retrieving automatically aggregated report on company's performance and statuses, we anticipate growth of the platform driven by User recommendations.

Furthermore, as domain experts in the fields, these investors and consulting firms will serve as important stakeholders during the early phase for us to improve the system based on their feedback.

After establishing a value proposition with the free version, we will introduce three pricing categories.

1. **Subscription Model:** The platform will be offered on a subscription basis, with tiered pricing depending on the size of the consulting firm and the level of functionality required.
2. **Pay-Per-Report:** Customers will be allowed to purchase individual reports on specific investments or markets, suitable for smaller firms, individual investors or less frequent users.
3. **Licensing:** The technology will be allowed to be licensed to larger firms or financial institutions that prefer to integrate it into their own systems.

2.6.2 Corporate Entities

We aim to provide value-adding services to Corporate Entities so that the companies could gain insights of the key metrics investors assessed to improve on their company's standings and performance. This would be in the form of reports or workshops where we will share data insights of top performing

companies in the same industry and derived key metrics that can improve the company further.

3. SYSTEM DESIGN AND MODEL

3.1 Architecture Overview

The system provides a user-friendly interface to streamline the process of gathering current and comprehensive information on the company of interest. The user can either key in the inputs to toggle the parameter (User Input) or choose to interact with the chatbot to scrape for necessary information with ease. It will either be consolidated on the webpage to compile the information in an orderly manner or be answered through the chatbot if one wants to find the information quickly without scoping through the whole webpage. This interface is designed in a way to offer a streamlined platform for comprehensive information gathering, provide actionable insights for informed decision-making, and establish benchmarks for performance assessment and comparative analysis between the company of interest and its competitors.

To automate the extraction of essential company background details and key financial metrics, TagUI was leveraged as the Robotic Process Automation (RPA) tool. This allowed the efficient retrieval of necessary information for the conduct of both technical and fundamental analyses, facilitating a comprehensive evaluation of the company's investment prospects.

The Real-Time News Informer Module is designed to systematically collect and analyze the latest news articles related to specified companies. Leveraging both an API and Robotic Process Automation (RPA), the module efficiently gathers news data from diverse sources, including NewsAPI for comprehensive coverage and Google News for targeted scraping. To ensure data integrity, duplicate removal techniques are employed, utilizing TF-IDF vectorization and cosine similarity measures. Sentiment analysis, powered by the Natural Language Toolkit (NLTK) with VADER lexicon, categorizes sentiments into positive, negative, or neutral, providing investors with actionable insights into market sentiment trends.

Following the collection and analysis of news articles, the module constructs a comprehensive sentiment summary by aggregating sentiment scores and calculating a weighted average score. This summary aids investors in quickly grasping prevailing market sentiment trends, facilitating faster and more informed decision-making.

The ESG system's architecture is designed to seamlessly integrate the collection, analysis, and reporting of ESG data for companies. At its core, the system

leverages a combination of RPA and AI to automatically gather data from various sources using `fetch_articles.py`, which employs the EventRegistry API for real-time news and report scraping. This architecture supports a modular analysis approach, with separate components handling environmental (`analyze_environmental.py`), social (`analyze_social.py`), and governance (`analyze_governance.py`) data using advanced NLP models from the transformers library. The results are then synthesized to calculate an overall ESG score, presenting a holistic view of a company's ESG profile.

Competitor analysis is performed utilizing Glassdoor reviews as a primary source of evaluation for the company's work culture. Initially, the top N competitors are identified through web scraping using TagUI via Google search. Following this, TagUI automates the process of logging into Glassdoor and extracting key metrics such as the employer overview rating, recommendation to friend rating, and CEO approval rating for the company of interest. Additionally, qualitative reviews are mined to extract the top 5 pros and cons, providing valuable insights into the overall sentiment of working for the company. Furthermore, ratings of the top N competitors are collected to establish a benchmark for comparison and ranking, facilitating a comprehensive assessment of the company's standings relative to its competitors.

3.2 System Overview / Detailed Workflow

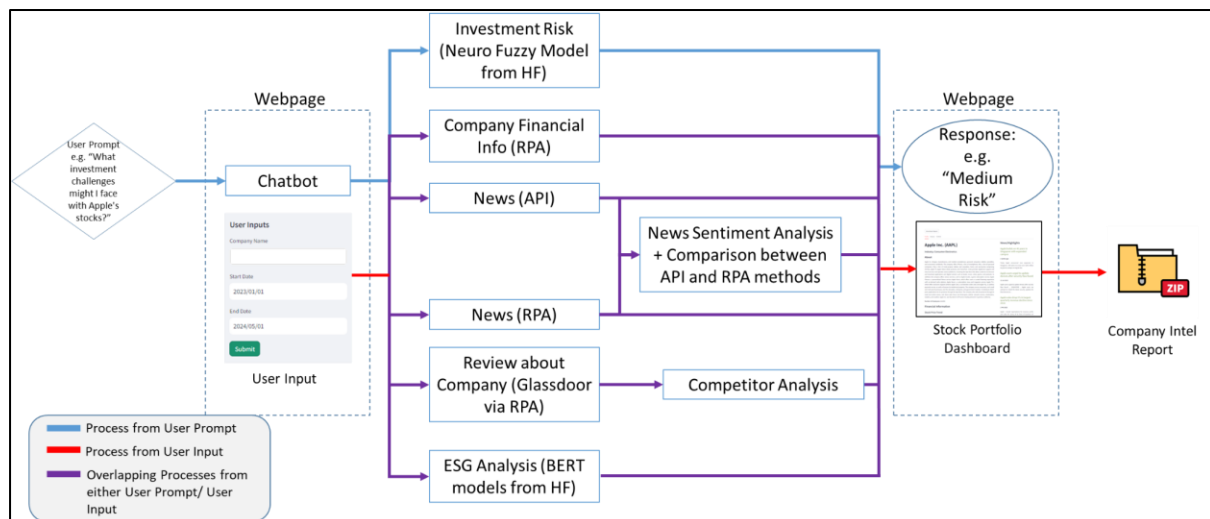


Figure 1: System Overview / Project Workflow

Above shows the flowchart of the project. Blue line represents the process of the chatbot, red line represents the process coming from the User Input and the purple means and overlap between the blue and red, which otherwise also means some processes run the same functions.

The inputs from the user will either be a User Input which contains parameters like company name, start date and end date, or the User Prompt which is under the Chatbot tab.

The output consists of Investment Risk of the company (which is only accessible via the chatbot), scrapped company's financial information, news, company's reviews via glassdoor and ESG analysis. User can choose to collect all the data by downloading the report which comes in the form of a zipped folder.

3.3 Technical Stack

3.3.1 Frontend:

- **Technologies:** Streamlit
- **Interface:** User interfaces are presented as web pages utilizing Streamlit, an open-source Python framework that is suitable for machine learning or data science practitioners.
- **Interaction:** The frontend communicates with the backend via functions imported from other Python files.

3.3.2 Backend:

- **Technologies:** Python
- **NLP Engine:** Processes user inputs and manages chatbot interactions.
- **ANFIS Neuro Fuzzy Model:** A deep learning model pre-trained on 3 years of stock data and derived features, and provides the investment risk based on past data statistics and recent stock data.
- **Web Scrapping:** Using python packages like TagUI and yfinance to scrape for financial statistics, news and company reviews when called upon.

3.3.3 Database:

- **Technologies:** SQLite
- **Purpose:** Store training dataset for neuro fuzzy model

3.4 Data Management

3.4.1 Data Sources

yahoo finance: Python package yfinance for fetching stock price data.

Features extracted: Open, Close, High, Low, Volume, Dividends

3.4.2 Data Preparation

Derived features for generating labels using auto encoder clustering in the neuro fuzzy model training:

Feature	Interpretation
Price_Range	Represents the daily high and low difference, giving a sense of the intraday volatility or the range within which the stock price moved.
Pct_Change	The daily percentage change shows how much the price has moved compared to the previous day, capturing the momentum.
SMA_7	SMA (Simple Moving Average over 7 days) smooths out price data by creating a constantly updated average price, which can help in identifying the trend direction.
EMA_7	EMA (Exponential Moving Average over 7 days) gives more weight to recent prices, which helps it react faster to price changes than SMA.
Volatility	Measured as the standard deviation of daily price changes, it indicates how drastically stock changes within timeframe
RSI	Relative Strength Index measures the speed and change of price movements and is typically used to identify overbought or oversold conditions.
MACD and MACD Signal	The Moving Average Convergence Divergence is a trend-following momentum indicator that shows the relationship between two moving averages of a stock's prices.
Bollinger Bands	These are volatility bands placed above and below a moving average. Volatility is based on the standard deviation, which changes as volatility increases or decreases.
VWAP	Volume Weighted Average Price gives the average price a stock has traded throughout the day, based on both volume and price. It is important as it provides insight into both the trend and value of a security.
RiskAndAdjustedReturn	Calculated by dividing the percentage change by the volatility. It measures the return earned per unit of risk taken.
IndustryRankRSI	Ranks the RSI values within an industry, giving a percentile-based position relative to peers.
CloseToVWAPRatio	Compares the closing price with the VWAP of the day.

MACDSignalDiff	The difference between the MACD and its signal line, highlighting crossovers that can indicate potential buy or sell signals
Percentage_Bandwidth	The width of the Bollinger Bands relative to the middle band, indicating the degree of market volatility.

4. SYSTEM DEVELOPMENT AND IMPLEMENTATION

4.1 Tools / Techniques

4.1.1 Gemma-2B Supervised Fine Tuning

Overview:

Supervised Fine Tuning (SFT) method:

Definition and process

Supervised Fine Tuning (SFT) is a method used to adapt pre-trained Large Language Models (LLMs) to specific tasks or datasets while leveraging their extensive initial training. The process involves taking a model that has been pre-trained on a vast corpus of data and further training it (fine-tuning) on a smaller, task-specific dataset under supervision. The "supervised" aspect refers to the use of labeled data, where the correct outputs are provided, allowing the model to learn the nuances of the new task more effectively.

Training Dataset Generation

The training dataset was crafted to include several key fields essential for effective conversation tracking and response formulation:

- **Conversation ID (conv_id):** This identifier is used to track individual conversations, enabling systematic review and analysis.
- **Step:** This metadata field captures the type of message within a conversation, helping to categorize user queries.
- **Question:** Represents the expected user query, facilitating targeted model training to handle specific user inputs.
- **Response:** Details the model's specific response to the user query, showcasing the system's ability to generate relevant replies.

Response Format

The format for the model's responses is designed to efficiently integrate various subsystems and manage their arguments. It is structured as follows:

- Format: **subsystem_name:args** (e.g., **stock_risk:Apple**)
 - **subsystem_name:** Indicates the function or subsystem being called.
 - **args:** Lists the arguments needed by the subsystem to execute its function.

Rationale for Response Format

This response format was selected to optimize the model's utility in real-world applications, specifically for its ability to:

- **Intent Recognition:** Identify the user's intent with precision.
- **Slot Detection:** Accurately detect and extract relevant information from the user query.
- **Subsystem Integration:** Seamlessly call the appropriate subsystem based on the identified intent and extracted slots, thereby facilitating a cohesive and functional response mechanism.

The chosen structure maximizes the model's effectiveness in handling diverse queries by integrating various functional subsystems based on the detected user intent and other contextual data.

Model Architecture:

```

Model: "GemmaForCausalLM"
-----
Layer (type)                Output Shape                Param #
-----
embed_tokens (Embedding)    (None, 256000, 3072)      786432000
-----
decoder_layers (Sequential) (None, 256000, 3072)      0
|
|-----
| Layer (type)                Output Shape                Param #
|-----
| gemma_decoder_layer_0 (GemmaDecoderLayer)
| |-----
| | self_attn (GemmaSdpaAttention)
| | |-----
| | | q_proj (Linear4bit)      (None, 4096)              12582912
| | | k_proj (Linear4bit)      (None, 4096)              12582912
| | | v_proj (Linear4bit)      (None, 4096)              12582912
| | | o_proj (Linear4bit)      (None, 3072)              12582912
| | | rotary_emb (GemmaRotaryEmbedding) - Dynamic shapes
| | |-----
| | mlp (GemmaMLP)
| | |-----
| | | gate_proj (Linear4bit)    (None, 24576)             75497472
| | | up_proj (Linear4bit)      (None, 24576)             75497472
| | | down_proj (Linear4bit)    (None, 3072)              75497472
| | | act_fn (GELUActivation)
| | |-----
| | input_layernorm (GemmaRMSNorm)
| | post_attention_layernorm (GemmaRMSNorm)
| |-----
| norm (GemmaRMSNorm)        (None, 256000, 3072)      0
-----
lm_head (Linear)            (None, 256000, 256000)    786432000
=====
Total params: 2,145,531,136
Trainable params: 2,145,531,136
Non-trainable params: 0
-----

```

Main Components:

The main components of the model include an Embedding Layer and several Decoder Layers, each equipped with specialized mechanisms to enhance language processing. The Embedding Layer, with dimensions (256000, 3072, padding_idx=0), is crucial for transforming input token indices into dense vector representations. This layer supports a large vocabulary of 256,000 tokens and provides high-dimensional embeddings (3072 features per token), enabling the model to handle a wide variety of language inputs with nuanced representations. The model's architecture features 28 stacked GemmaDecoderLayer instances, each designed to process language inputs sequentially. This setup allows the model to generate language outputs one token at a time, influenced by previously generated tokens.

At the heart of each decoder layer is the GemmaSdpaAttention mechanism, which includes 4-bit linear projections for queries, keys, values, and outputs. These projections maintain computational efficiency by reducing precision to 4-bit. Additionally, the GemmaRotaryEmbedding enhances positional encoding, improving the model's understanding of word order and relationships in a sequence. Each decoder also incorporates a GemmaMLP, featuring Gate and Up/Down Projections for dimensional adjustments and a GELU activation function to add non-linearity, crucial for modeling complex text patterns. Two RMS normalization layers (GemmaRMSNorm) per decoder layer ensure stable training dynamics and effective contributions, one before and one after the attention mechanism.

The output layer, known as lm_head, maps the decoder's output back to the vocabulary size with dimensions (3072, 256000), facilitating the prediction of the next token in the sequence. The GemmaForCausalLM model offers several advantages, including efficiency through the use of 4-bit linear projections, which significantly reduces memory usage and computational overhead. Despite the reduced precision, the model maintains high performance in language understanding and generation. With its 28 layers, the model exhibits flexibility and adaptability to complex language patterns, making it suitable for advanced tasks like story generation and dialogue systems.

4.1.1 Assessing Key Company Information and Financial Metrics

Overview:

In the investment decision-making process, it's essential to consider various factors such as company background, financial metrics, technical and fundamental analyses, and the company's work culture. To facilitate this comprehensive evaluation, we employed a robotic process automation tool, to compile real-time data and analyze these critical pieces of information.

Choice of RPA Tool

Besides TagUI, BeautifulSoup and yfinance python packages were considered and examined for use as the choice of RPA tool to scrape the real-time data.

However, BeautifulSoup faced limitations in navigating across various tabs within a webpage or transitioning to different webpages to gather the diverse information. Due to its constraint in scraping information solely from predefined webpages, it became impractical to specify unique URLs for each company of interest. Consequently, BeautifulSoup proved less effective in our context, especially when dealing with websites where URL links were unknown or varied.

The yfinance library was effective in collecting the real-time financial market data (up to 1-minute intervals) from Yahoo! Finance without the need of accessing webpages. To validate accuracy, a comparison code was developed to compare the results from real-time scrapping using TagUI and yfinance, enabling regular validation in future. The results showed reasonable alignment, with minor discrepancies observed only during live trading sessions. A limitation however is the use of ticker symbols to retrieve the market data. This may not be practical for the User to input and would be a challenge to create a database library upfront which stores all the companies' corresponding ticker symbols across all stock exchange markets worldwide for conversion. Hence, TagUI is still preferred for end-to-end scrapping of information.

Scraping Company Information and Financial Metrics:

Background information on the Company was extracted to provide users with insights into the Company's profile, business operations and key personnel. Additionally, key financial metrics scrapped offers users an understanding of the company's performance and financial health. The list of information and metrics extracted are as follows.

<u>Company Information</u>	<u>Market Performance Metrics</u>
Full company name	Live price
Ticker symbol	Previous close
Industry	Opening price
Employee count	Day price range (low and high)
Company overview	Market capitalization
Key personnel	Earnings per share (TTM)

<u>Trading Metrics</u>	<u>Financial Ratios</u>
Trade daily volume	Price-to-sales ratio (TTM)
Currency	Price-to-book ratio
Stock market	Price-to-earnings ratio (TTM)
	Return on equity (ROE)
<u>Dividend Metrics</u>	Return on assets (ROA)
Forward annual dividend rate	Net profit margin
Forward annual dividend yield	Debt-to-equity ratio
Trailing annual dividend rate	Current ratio
Trailing annual dividend yield	

Technical Analysis:

The financial ratios were further utilized to evaluate the financial health and performance of the company. Generic benchmark is used to return a rating of “poor”, “fair”, “good” or “excellent” rating for each of these ratios. An overall scoring was then used to demonstrate how recommendation to buy, hold or sell could be determined, where each of the rating is assigned a score and having buy and sell thresholds established. These benchmarks and thresholds could be further refined based on domain expert’s advice subsequently to develop a more accurate investment recommendation.

Fundamental Analysis:

As part of the fundamental analysis, an in-depth examination of stock price trends was conducted using line chart snapshots captured at various intervals via TagUI. To begin the analysis, these images underwent preprocessing with OpenCV, a process designed to isolate the plot area and extract the essential features for interpreting stock price movements over selected durations. This preprocessing step ensured that only the relevant data points were considered in subsequent analyses.

Following preprocessing, the coordinates of the detected edge lines were extracted and utilized to segment the chart into distinct quarters, facilitating a granular assessment of stock price movements over time. This quarter-wise breakdown provided valuable insights into the evolution of the stock's performance throughout the selected period.

The most recent quarter served as a focal point for generating relevant descriptions, offering insights into whether the latest stock price exhibited

improvement or deterioration compared to the beginning of the quarter. This comparative analysis allowed for a nuanced understanding of the stock's recent trajectory and its potential implications for investors.

Moreover, beyond the basic assessment of stock price movements, further insights and interpretations can still be gathered from the charts through collaboration with domain experts. For instance, these experts can provide insights to discern subtle indicators, such as the slope of the detected edge lines, which may indicate whether the stock has exhibited signs of stabilization, consolidation, market correction, recovery, and more. By integrating such analyses, investors can derive more meaningful outcomes, empowering them to make informed decisions regarding their positions with respect to the stock.

Glassdoor Review:

To evaluate the company's work culture, Glassdoor reviews served as a valuable resource. The top N competitors were firstly scrapped using TagUI via google search. Subsequently, TagUI proceeded to log into Glassdoor and search for the company of interest to extract key metrics of the employer overview rating, recommendation to friend rating for working in the company and the CEO approval rating. Additionally, qualitative reviews were mined to extract the top 5 pros and cons, offering insights into the general sentiment of working for the company. Lastly, ratings of the top N competitors were collected to provide a benchmark for comparison and ranking, enabling a comprehensive assessment of the company's standings.

4.1.2 ESG Integration in Financial Analytics

The ESG system enhances traditional financial metrics analysis by integrating a comprehensive ESG perspective, providing a multi-dimensional view of a company's operational performance and resilience. Utilizing scripts like `analyze_environmental.py`, `analyze_social.py`, and `analyze_governance.py`, the system meticulously evaluates a company's sustainability initiatives, social impact, and governance practices alongside financial data such as profitability ratios, liquidity ratios, and market valuation.

In particular, the environmental module evaluates metrics related to carbon footprint, resource management, and compliance with environmental regulations. The social module examines factors such as community engagement, employee relations, and diversity practices. The governance module assesses board composition, policy adherence, and ethical conduct.

This integration allows investors to not only gauge a company's financial health but also its long-term sustainability and ethical alignment, which are increasingly

important in today's market. The system processes this data to extract actionable insights, which are then presented through a user-friendly dashboard interface, enabling investors to make well-informed decisions aligned with modern ESG standards.

The ESG system heralds a transformative approach to financial analysis by embedding Environmental, Social, and Governance (ESG) considerations directly into the core assessment of company profiles. Leveraging the power of specialized BERT models—EnvRoBERTa for environmental, SocialBERT for social, and GovRoBERTa for governance factors—the system performs nuanced textual analysis, drawing insights from vast quantities of unstructured data with a degree of accuracy and depth that traditional models cannot match. These models, sourced from Hugging Face's Model Hub, have been fine-tuned on ESG-focused datasets to ensure relevance and precision in scoring.

Each factor-specific BERT model is utilized separately to provide a focused analysis tailored to the nuances of environmental, social, and governance factors. This approach enables a deeper understanding of each aspect of a company's ESG performance, allowing for more granular insights and actionable recommendations. The resulting scores are derived from a combination of textual analysis and contextual understanding, incorporating news articles, financial reports, and other relevant sources of information. This comprehensive approach to scoring ensures that the ESG assessment is not only accurate but also robust and reliable.

Through advanced natural language processing (NLP), each BERT model processes related textual data, rating the company's performance across ESG dimensions. The resulting scores are presented through three distinct approaches, catering to different investor needs and perspectives. The flexibility in score presentation ensures that there is no one-size-fits-all rating, empowering users to interpret and utilize the scores according to their investment philosophy and criteria.

As the ESG system matures, the front-end implementation will evolve from Flask to Streamlit, a decision driven by the development team's Python-centric expertise. This transition will facilitate an even more seamless and interactive user experience while maintaining the robust analytical capabilities of the back-end.

4.1.3 Real-Time News Informer Module

News Scraper & Analysis Module Overview

The necessity for timely and precise information in investment decision-making has led to the development of a real-time news scraping and analysis module. This module is designed to systematically collect and analyze the latest news articles related to specified companies from various sources, addressing the challenges of data volume and velocity that investors face daily. Utilizing an API and RPA, the system efficiently gathers news, providing it to a processing pipeline to evaluate the collective sentiment, generating a distilled summary that aids in swift decision-making.

The module aggregates and analyzes news articles published within the last 48 hours, focusing on the most recent and relevant information. By harnessing NewsAPI for wide-reaching news collection and Google News through robotic process automation for targeted scraping, the system ensures comprehensive coverage. The combined use of these tools facilitates the calculation of sentiment scores across various articles. These scores are synthesized to present a consolidated sentiment summary, enabling investors to gauge market sentiment promptly and effectively.

Integration of NewsAPI for Comprehensive Data Collection

NewsAPI is integral to the module, offering access to a broad spectrum of news sources that span local and international media outlets. It functions by retrieving JSON-formatted search results for news articles using specific queries related to the companies of interest. This API is especially valuable for its capacity to pull both current and historical news data from over 150,000 sources globally, ensuring that no critical piece of information is missed in the analysis. This wide reach is crucial for creating a complete picture of the news landscape surrounding each specified company, providing a foundational layer for the sentiment analysis.

Robotic Process Automation for Dynamic News Scraping

To complement the breadth of NewsAPI, the system employs RPA to specifically target and scrape news from Google News. This approach was adopted after initial trials with BeautifulSoup proved ineffective due to the static nature of URLs in Google News searches. Instead, TagUI was implemented to automate interactions with the web interface, enabling the module to navigate through pages, apply custom filters, and extract news titles and summaries effectively. This method allows for the precise collection of data from Google News, ensuring

that the dataset includes the most current news items, which are pivotal for real-time analysis.

Enhancing Data Integrity Through Duplicate Removal

To maintain the accuracy and relevance of the data collected, the system incorporates a sophisticated duplicate removal process. This is achieved using the Term Frequency-Inverse Document Frequency (TF-IDF) vectorization combined with cosine similarity measures to identify and eliminate nearly identical articles. We set a high similarity threshold of 0.9 to ensure that only genuinely duplicated content is filtered out. This step is crucial as it prevents the redundancy that can skew sentiment analysis results and ensures that each piece of news is unique and contributes meaningfully to the overall sentiment assessment. This methodical approach to data cleansing is vital for maintaining the quality of the analysis provided to the end users.

Segmented Sentiment Analysis

Following the collection and cleansing of news articles, the system performs sentiment analysis to classify the emotional tone of each piece. We utilize the Natural Language Toolkit (NLTK) with its VADER lexicon, which is specially designed for sentiment analysis of social media texts but is highly effective for general news content as well. VADER considers the valence of words and phrases, providing a composite score that categorizes sentiments into positive, negative, or neutral. Each article's sentiment score is categorized as positive if the article sentiment score is 0.52 or higher, negative if less than -0.48, and otherwise neutral to maintain strong confidence in the outputs given. This sentiment analysis not only quantifies the sentiment of individual articles but also aggregates these results to produce a holistic view of the sentiment surrounding a specified company. This step is instrumental in synthesizing complex data into actionable insights, which is critical for investors making time-sensitive decisions.

Sentiment Summarization and Interpretation

Once sentiment analysis is complete, the system aggregates the data to construct a comprehensive sentiment summary. This involves summarizing the counts of positive, negative, and neutral articles derived from both NewsAPI and RPA sources. These counts are crucial for understanding the broader sentiment trends across different. Additionally, the average sentiment scores from each source are calculated and then combined into a weighted average that reflects the overall sentiment more accurately. This weighted score is critical for providing a nuanced

understanding of the market mood as it incorporates the variances in sentiment distribution among different news sources.

The final step in our analysis involves interpreting the aggregated sentiment to provide a straightforward market sentiment summary. This is done by analyzing the combined average sentiment score, where the system classifies the overall sentiment. This interpretation helps users quickly grasp the prevailing sentiment trend without needing to delve into detailed data, thus facilitating faster and more informed decision-making. By providing this summary, the module aids investors in understanding the general market sentiment at a glance, which can be crucial during volatile market conditions.

4.1.4 Neuro Fuzzy Model

Overview:

The provided Adaptive Neuro-Fuzzy Inference System (ANFIS) model blends neural network methodologies with fuzzy logic principles to create a flexible system that can effectively model complex nonlinear functions. This implementation is designed for predicting stock risk for companies, making it a valuable tool in financial analytics. Here's a detailed breakdown of its components and functionalities.

Components of the ANFIS Model

The Adaptive Neuro-Fuzzy Inference System (ANFIS) model integrates multiple layers to transform crisp input values into a probabilistic output that categorizes risk. The process begins with the Fuzzification Layer, where crisp inputs are converted into fuzzy values using predefined fuzzy sets and Gaussian membership functions. This layer's functionality is defined by several parameters including the number of input variables (`num_variables`), the number of fuzzy sets per variable (`num_sets`), and the parameters for the Gaussian functions (means and sigmas), which determine the centers and widths of the fuzzy sets. These parameters are learned during training to optimize performance.

Following the initial fuzzification, the Rule Application Layer takes over, utilizing the fuzzy values to determine the firing strength of each fuzzy rule. This is achieved through fuzzy logic operations. The layer is characterized by the number of output classes (`num_classes`), the reinforcement of the multi-dimensional input space through the number of input variables (`num_variables`), and the total number of fuzzy rules (`num_rules`). A specialized dense neural layer with a sigmoid activation function (`rule_combination_layer`) combines the fuzzified inputs to compute the rule firing strengths.

The process culminates in the Defuzzification Layer, where the fuzzy rule outputs are synthesized into a crisp output value. This layer uses parameters such as the number of rules (num_rules) and the number of output classes (num_classes), alongside trained rule weights, which quantify each rule's contribution to the final class scores.

The ANFIS model encapsulates these layers into a cohesive structure capable of performing end-to-end processing from raw inputs to predicted output classes. The model operates in three main stages: initially, it processes raw input data (e.g., financial indicators) through the fuzzification layer to generate fuzzy values. These fuzzified values are then input into the rule application layer, where the firing strength of each fuzzy rule is determined. Finally, the defuzzification layer aggregates these rule outputs to produce a probability distribution over potential risk categories, utilizing a softmax function. This integrated framework allows for the simultaneous training of the fuzzification, rule application, and defuzzification components through backpropagation, enhancing the model's efficiency and effectiveness in handling complex input-output relationships.

4.1.5 Backend

This system's backend relies on Python as its core programming language. A Natural Language Processing (NLP) engine is responsible for understanding user inputs and orchestrating the chatbot's conversational interactions. At the heart of investment risk analysis lies an ANFIS Neuro-Fuzzy Model. This deep learning model leverages a 3-year dataset of historical stock data and calculated features to provide risk assessments informed by both past performance and current market trends. Python libraries including TagUI and yfinance enable web scraping, empowering the system to gather real-time financial statistics, news updates, and company reviews on demand.

SQLite serves as the database technology. The primary purpose of this database is to store the essential training dataset used to refine the ANFIS Neuro-Fuzzy Model.

4.1.6 Frontend and Others

Frontend

Streamlit is a framework that has been utilized to design the frontend of the project. Streamlit is being chosen because of its simplicity, which means it is fairly easy to deploy. The webpage itself is made up of 3 tabs, which are 'Profile', 'Analysis' and 'Chatbot'. The user can start off by either keying in the parameters

under the ‘User Input’, or interact with the chatbot under the ‘Chatbot’ tab to get answers instantly. The ‘Download Report’ button will provide a copy of the information displayed on the web page to the user after querying. Below explains the functionality of the 3 tabs.

Profile

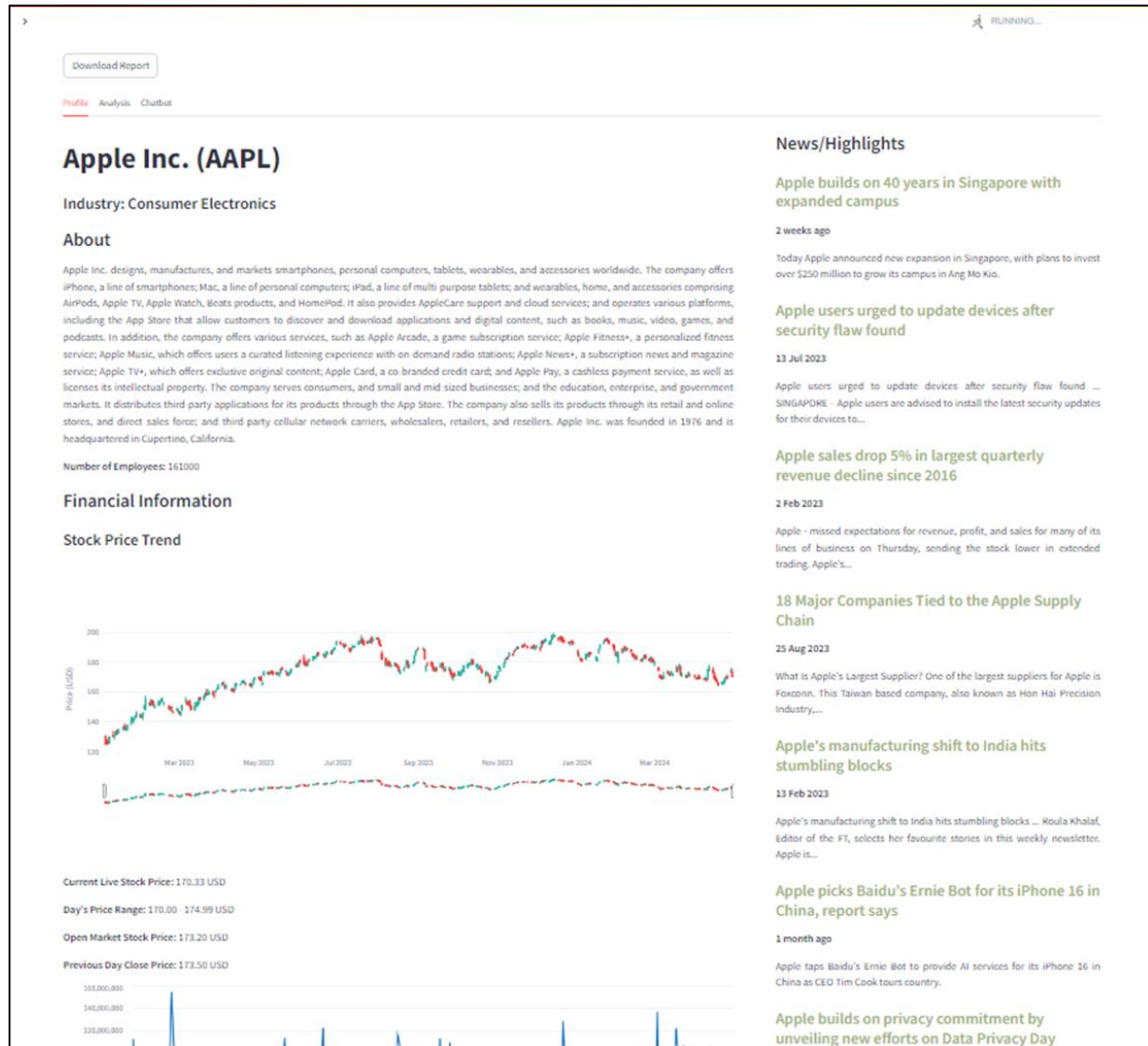


Figure 2: Screenshot of Profile Tab

This tab is supposed to contain facts like financial statistics, latest news and highlights, key personnel as well as the context of the company the user is interested in. This tab is supposed to serve as a one-stop hub for the user to understand the interested company as much as possible.

Analysis



Figure 3: Screenshot of Analysis Tab

Data that are revealed here have gone through a certain amount of analysis. If the user wants a deeper understanding of the company, he or she can look into this tab. Financial ratios like Price to Sales ratio, Return on Equity, and more are shown here to give users a better understanding to sell or hold. News of the company has gone through sentiment analysis to see the reputation of the company during the indicated period of time. In addition, employees' review about the company is also compiled here to provide a non-bias view. Moreover, the international standing of the company can be seen as attributes like company's rating by employees, company's job recommendation as well as CEO approval are being compared with its competitors. Last but not least, predicted ESG ratings of the company are presented to reveal the strength of sustainable and ethical practices, since they impact the company's long-term resilience and performance in the long term.

Chatbot

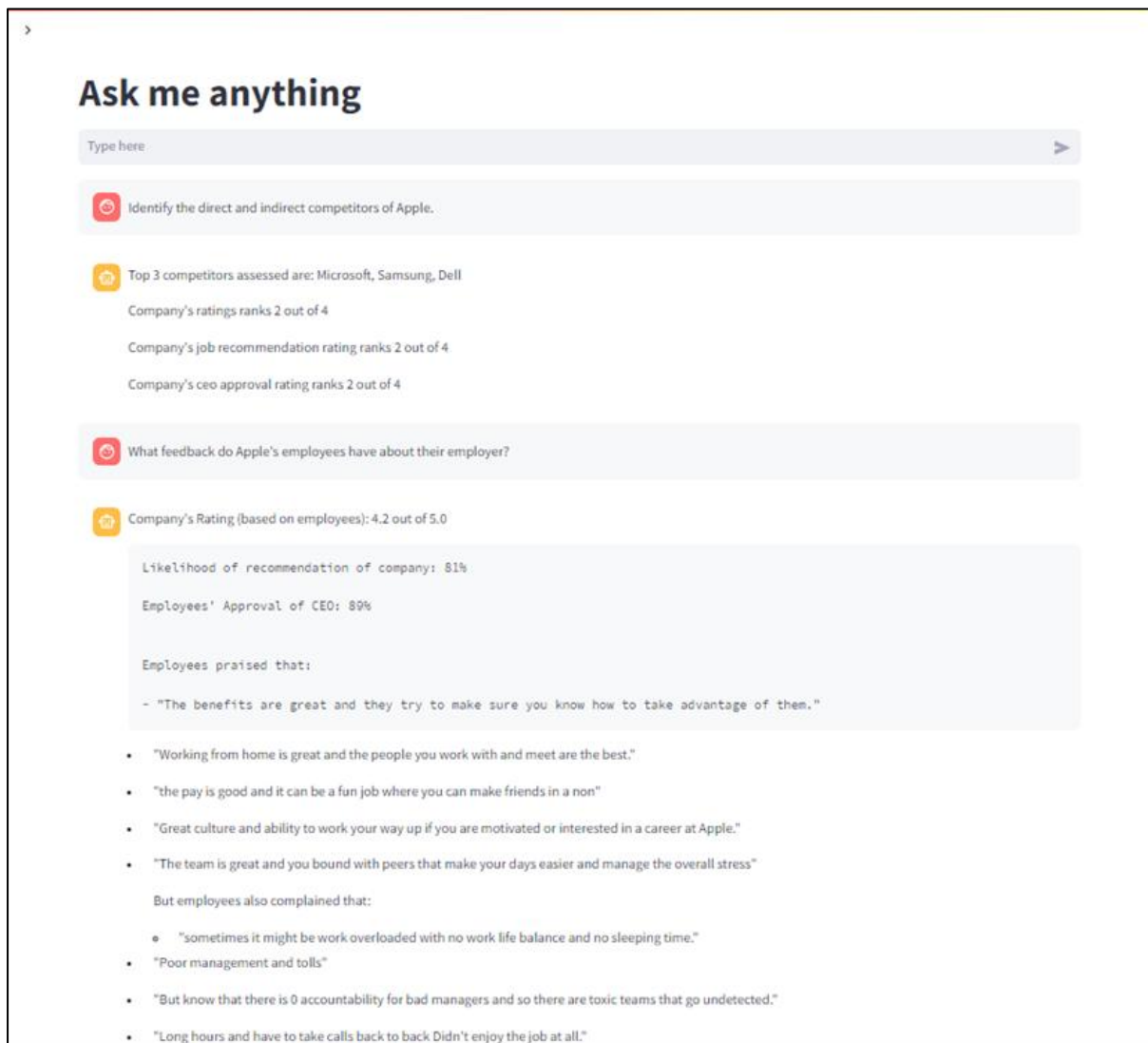


Figure 4: Screenshot of Chatbot

The Chatbot tab can tend to the queries of the user if the user has a specific question in mind or if he has trouble finding the relevant information on the webpage.

1. If the user were to ask questions about risk in investing in a certain company, the chatbot replies with 'Low Risk', 'Medium Risk' or 'High Risk'. This is depending on the data that is trained on GemmaForCausalLM.
2. If the user were to ask questions about a company's financial standing, the chatbot can scrape for the information for within the specific duration tasked.

3. If the user were to ask questions about the company's competitors, the chatbot is capable of revealing the competitors of the company and run a comparison amongst the competitors in terms of employees' ratings, CEO's approval as well as job recommendation of the company.
4. If the user were to ask questions about employees' reviews about the company, the chatbot can scrape for summarized reviews done by employees of the company.
5. If the user were to ask questions with regards to the latest happenings of the company, news or latest highlights will be scraped and shown to users.

Chat history is being saved and can be downloaded from the report.

4.2 Challenges and Solutions

4.2.1 Nature of Streamlit framework

The downside of Streamlit is when the web page is being interacted with, the script will rerun, which might cause some information to be wiped from the display. To resolve this, historical data is being saved into files, in addition to the implementation of conditions, Streamlit will always load the historical files whenever the webpage is interacted to show as if the webpage is static.

4.2.2 Parsing User Input for the NLP Chatbot

Chatbot is not trained to augment the answer retrieved from appropriate functions, so when the answer is being thrown back to the user, it does not provide an immersive experience such as another human is conversing to the user. Answers provided back all come in a template form, hence the query of the user only fetches for the answer which could be the closest to what the user needs, but might not be the actual answer. In future, an additional model like BART which is a considerably small model, can be implemented to augment the answer before throwing it back to the user.

5. FINDINGS AND DISCUSSION

5.1 Risk Analysis

5.1.1 Technical Risks

One of the fundamental concerns is system reliability and uptime. Given the platform's reliance on real-time data, any interruptions in service could significantly disrupt user operations. To mitigate this risk, the implementation of regular stress testing and robust redundancy plans is crucial. These measures will help identify potential failure points in the system infrastructure, allowing for the development of effective fallback procedures and ensuring continuity of service during unexpected downtimes.

Another critical aspect is data security and privacy. As the platform handles sensitive financial data, it must incorporate stringent security measures to prevent data breaches and ensure compliance with data protection regulations such as the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA). This involves encrypting data both in transit and at rest, implementing strict access controls, and regularly updating security protocols to guard against emerging cybersecurity threats.

Finally, scalability poses a significant technical challenge. The platform must be capable of handling varying loads, scaling effectively during peak usage times without degradation in performance. This requires ongoing investment in scalable cloud services and architectural solutions that support elastic scaling. Efficient resource management strategies, such as load balancing and elastic computing resources, will be essential to accommodate growth in user numbers and data volume without compromising the speed or functionality of the analytics provided.

5.1.2 Compliance Risks

In the dynamic and stringently regulated financial sector, the Investor Intelligence Agent must meticulously adhere to international finance regulations. These regulations, encompassing a broad spectrum from data protection laws such as GDPR in Europe to financial reporting standards like the Sarbanes-Oxley Act in the United States, can vary significantly across different jurisdictions and are subject to frequent revisions. This necessitates a proactive regulatory compliance strategy, which includes regular updates to the system in response to new legal requirements and ongoing staff training to ensure all regulatory changes are effectively understood and implemented.

Additionally, the platform faces substantial data governance challenges. High-quality data management is crucial not only for operational excellence but also to

comply with stringent financial reporting and privacy standards globally. To address this, the platform must establish and maintain a sophisticated data governance framework. This framework should include mechanisms for data quality assurance, effective data handling procedures, and comprehensive periodic audits to verify compliance with all applicable data protection laws and financial regulations. These audits help identify potential compliance risks early, allowing for timely adjustments in data management practices, thereby ensuring the platform remains compliant and trustworthy.

By embedding these compliance measures into the core operational strategy, the Investor Intelligence Agent aims to mitigate potential legal and financial penalties while safeguarding its reputation as a reliable and compliant financial analytics tool. This approach not only supports sustainable business growth but also builds trust with users by demonstrating a strong commitment to legal and ethical standards.

5.1.3 Market and Operational Risks

The adoption of the Investor Intelligence Agent in the target market presents significant risks. Despite the platform's advanced features and market positioning, there's a possibility that it may not achieve anticipated traction swiftly. To mitigate this risk, thorough market research and robust marketing strategies tailored to the specific needs and preferences of small to mid-sized investment firms and individual investors are crucial. These efforts should focus on demonstrating the tangible benefits of the platform and how it addresses specific pain points unique to these segments.

Moreover, the platform's reliance on third-party services for critical data feeds and computational resources introduces vulnerabilities. Any interruption or degradation in these services could impair functionality, leading to service downtimes or compromised data integrity. Establishing strong partnerships and having contingency plans, such as secondary data sources and computational backups, are essential strategies to ensure continuity and reliability of service.

Intellectual property risks also pose a substantial challenge. The use of third-party components within the platform necessitates diligent management to ensure all elements are properly licensed, and that the platform does not inadvertently infringe on existing patents or copyrights. This requires a proactive intellectual property audit strategy and ongoing monitoring to avoid legal entanglements that could not only result in significant financial liabilities but also damage the platform's reputation and market standing.

5.1.4 Financial and Reputational Risks

During the development and deployment of the project, two critical areas of risk—financial and reputational—warrant special attention. Financial sustainability hinges on securing adequate funding to support both the developmental and operational phases of the project. Insufficient capital can lead to delays in achieving project milestones, potentially stalling the platform's launch and scaling phases. This requires meticulous financial planning and proactive fundraising strategies to ensure a steady cash flow, thereby avoiding potential bottlenecks that could impede progress.

On the reputational front, as a newcomer in the competitive field of financial analytics, establishing credibility is paramount. The platform's success is heavily dependent on user trust, which can be swiftly undermined by failures in product performance or customer service. Early adopters will be particularly sensitive to these aspects; hence, ensuring that the platform is robust, reliable, and user-friendly from the outset is crucial. Additionally, effective communication and swift customer support are essential to mitigate any potential dissatisfaction among users. Addressing these risks involves not only rigorous testing and quality assurance before release but also a strong commitment to customer engagement and service post-launch.

5.1.5 ESG Analysis Risks

Integrating ESG analysis introduces specific risks that require careful consideration. Subjectivity and the absence of standardized metrics may result in varying interpretations and evaluations, potentially impacting the reliability of assessments. Additionally, the risk of "greenwashing" exists, where companies may manipulate ESG data to present a more favorable image. Ensuring data quality and reliability poses another challenge, given variations in data availability and consistency. Moreover, evolving regulatory requirements surrounding ESG disclosure may lead to compliance risks. Proactively managing these risks through transparent methodologies, rigorous validation processes, and adherence to regulatory standards is crucial for maintaining the integrity of ESG analysis.

5.2 Roadmap for Future

5.2.1 Future Considerations

As the system developed is positioned as an MVP, further enhancements listed below are recommended to improve the system.

Web Scraping via TagUI

Data scraped from one website and applied on the search engine of another website may not lead to expected results. For example, when names of Tesla's competitors on Google are used to search on Glassdoor, it may not lead to competitors' pages directly since they are spelled differently on Glassdoor instead. To optimize on this, exception code can be produced to scrape for different variations of the same data or to use an existing API to engage search engines in a more optimized manner.

ESG Analysis

As we look to the horizon of ESG analysis, the ESG system is envisioned as a forerunner in shaping the discourse on sustainable and ethical investment. It will readily assimilate ongoing disclosures and regulatory shifts within the ESG landscape, bolstered by the predictive prowess of evolving BERT models.

Future iterations of the ESG system will venture beyond the conventional bounds of ESG, exploring the intricate interplay between environmental factors and other, more latent dimensions of sustainability. The potential integration of machine learning and deep learning will further refine the analytical process, unveiling the subtle interconnections within ESG factors.

The anticipated integration with Streamlit will enhance the system's user interface, propelling it towards a more dynamic and engaging user experience. The anticipated features will not only offer real-time ESG tracking but also deliver customizable dashboards and interactive visualizations, shaping a personalized user journey.

Enhancing Competitor Analysis

In future iterations, competitor analysis can be further enhanced by incorporating additional financial metrics derived from fundamental and technical analysis, alongside sentiments extracted from news sources and Environmental, Social, and Governance (ESG) scores. The existing framework used for researching the company of interest could be adopted for its key competitors to provide a more holistic understanding of a company's competitive landscape. By integrating financial performance indicators such as revenue growth, profit margins, and stock price trends, analysts can gain deeper insights into the financial health and market positioning of both the company of interest and its competitors. By amalgamating these diverse data sources, future iterations of competitor analysis

can offer a more well-rounded perspective, enabling informed decision-making on evaluating the company of interest's performance.

Developing Benchmarks for Evaluating Companies

As the database expands, there will be opportunities to establish benchmarks that can serve as robust metrics for evaluating performance of companies. By leveraging data analytics and predictive analysis techniques, we can leverage this wealth of data to identify key areas and factors for improvement within a company's operations. The insights gleaned from these analyses can empower organisations to make informed decisions to allocate resources more effectively and prioritize initiatives that yield the greatest impact.

Integrated Predictive Modelling

The next significant step in enhancing the Investor Intelligence Agent involves the development of an Integrated Predictive Modelling feature, which will leverage advanced machine learning algorithms to forecast market trends and investment risks. This feature will utilize a combination of historical data, real-time market inputs, and predictive analytics to provide users with forward-looking insights into potential market movements. By integrating models such as ARIMA for time series forecasting and LSTM networks for capturing complex temporal patterns in financial data, the system will be able to offer predictions with higher accuracy. Additionally, the predictive model will be continuously refined and retrained with new data to adapt to changing market conditions, ensuring its relevance and efficacy. This feature aims to empower users by providing a proactive tool for making informed investment decisions, significantly enhancing their ability to anticipate market changes and adjust their strategies accordingly.

Dynamic Scenario Analysis Tool

Another future consideration for the product is a scenario planning platform. As scenario planning becomes increasingly relevant in the uncertain environment seen today, it will add value to financial planners' workflows. This tool will empower users to simulate and analyze various financial scenarios by manipulating key market variables such as interest rates or sector-specific changes. Leveraging historical data and advanced predictive analytics, users can model the potential impacts of different scenarios on their investments. The primary benefit of this tool is its capacity to enable proactive risk management and strategic planning. By visualizing the outcomes of hypothetical market conditions, users can make informed decisions that optimize their investment strategies, ensuring they are well-prepared for a range of possible future states.

Real-time Regulatory Change Tracker

To address the challenge of keeping up with the fast-evolving regulatory landscape in financial markets, we plan to integrate a Real-time Regulatory Change Tracker into the Investor Intelligence Agent. This tracker will continuously monitor and report on updates in financial regulations across various jurisdictions, utilizing cutting-edge natural language processing techniques to distill and summarize regulatory changes. The tool will alert users to new regulations and provide concise, actionable insights on how such changes might impact different markets or specific stocks. This feature is crucial for maintaining compliance and enables users to swiftly reposition their investment strategies in response to new regulatory requirements, thereby safeguarding their investments against potential legal and financial repercussions.

6. CONCLUSIONS

The minimal viable product (MVP) of the Investor Intelligence Agent has conclusively demonstrated its ability to meet and exceed the initially defined outcomes. These achievements are evidenced by its ability to deliver personalized financial insights through advanced AI integration, manage real-time data effectively for timely decision-making, and offer comprehensive ESG assessments that align with current investment criteria and regulatory standards.

The Investor Intelligence Agent project has uncovered a robust demand for AI-driven financial advisory tools that address real-time market analysis and comprehensive financial data integration. The extensive market research underscores the necessity for innovative solutions that can process vast amounts of information rapidly and accurately, catering to a dynamic financial landscape. Key technological innovations like the real-time news informer module and advanced ESG integration uniquely position the platform to meet these industry demands effectively.

The Investor Intelligence Agent is uniquely crafted to empower a wide range of users from individual investors to mid-sized investment firms, offering advanced tools that were traditionally accessible only to large institutions. The platform's core strength lies in its integration of sophisticated AI analytics and adaptive learning models, such as the Gemma-2B and ANFIS, which provide personalized investment insights and risk assessments. The system's use of RPA, particularly TagUI, enhances its ability to deliver timely and accurate data, making complex financial analysis more accessible and customizable to the specific needs of its users.

Looking ahead, the Investor Intelligence Agent is designed for scalability and continual enhancement. Future development plans include the integration of a Dynamic Scenario Analysis Tool and a Real-time Regulatory Change Tracker, which will further solidify the platform's capability to adapt to and anticipate market changes. These enhancements will leverage ongoing technological advancements to ensure the platform not only keeps pace with the industry but sets new standards for innovation and performance in financial analytics. The commitment to continuous improvement reflects the vision to lead in creating more intelligent, responsive, and user-focused financial tools.

APPENDIX A – INSTALLATION AND USER GUIDE

Installation

- Download repository folder and see <Instructions on how to run the Code> below to run the system.

Directory Layout

```
SystemCode/
|
|-- requirements_2.txt
|
|-- Frontend/
|   |-- main.py
|   |-- style.css
|   |-- Cover.jpg
|
|-- Backend/
|   |-- .Neuro_Fuzzy/
|   |-- .ESG/
|   |-- Company_Scrapping.py
|   |-- GD_SentimentAnalysis.py
|   |-- .history/
|   |-- constant_api_key.py
|
|-- Model_training/
|   |-- Clustering.ipynb
|   |-- generate_training_data_gemma.ipynb
|   |-- neuro_fuzzy.ipynb
|   |-- fine_tune_gemma.ipynb
```

Brief Description

- The main.py in the Frontend directory serves as both the main script and the webpage design. Interactions with user inputs or prompts to the chatbot trigger web scraping (Company_Scrapping.py, GD_SentimentAnalysis.py) and model inferencing (ESG_main.py, gemma_inference.py).

Instructions on how to run the Code

1. Navigate to the SystemCode directory:
`cd SystemCode/`
2. Install required packages

```
pip install -r requirements_2.txt
```

3. (Optional) Set up EventRegistry and NewsAPI respective tokens in Backend/.env
 - a. EventRegistry Setup: <https://eventregistry.org/>
 - b. NewsAPI Registration: <https://newsapi.org/register>
4. Run the Application

```
streamlit run Frontend/main.py
```

System Requirements

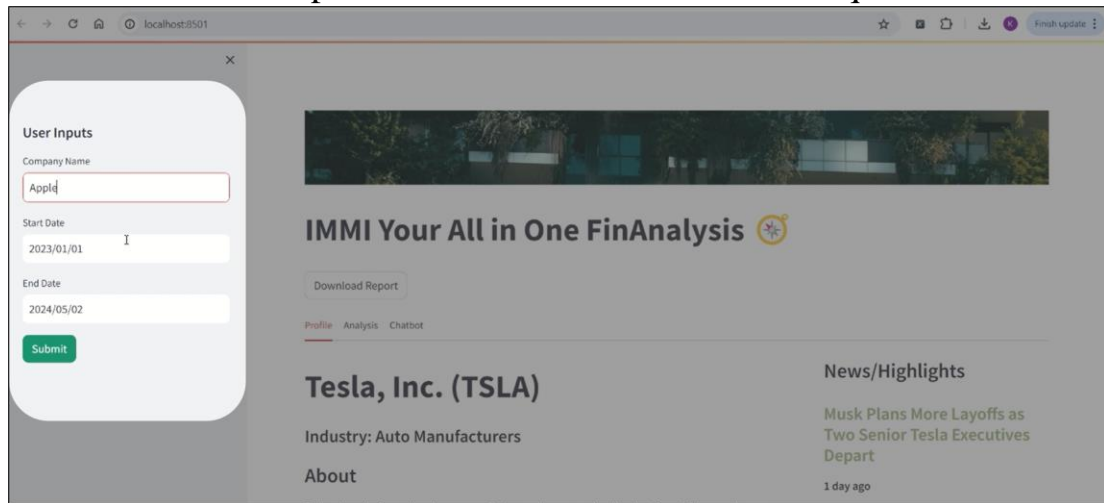
- Browser: Google Chrome
- RAM: Minimum of 32GB required to run the Gemma model

Additional Resources

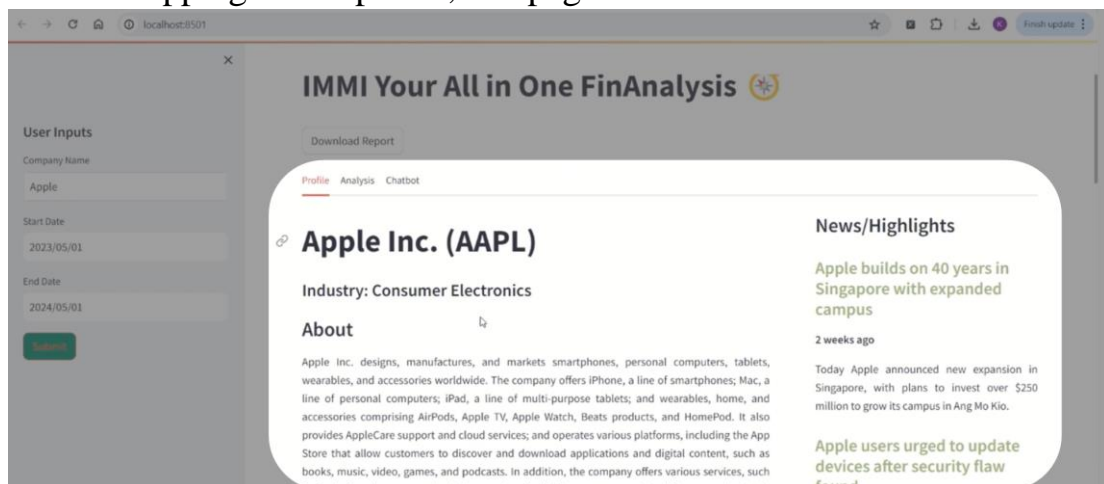
- Gemma Model: Check out Gemma model here:
<https://huggingface.co/yatharth97/yatharth-gemma-2b-it-isa-v2>
- Dataset: The dataset used to fine-tune the Gemma model is available here:
https://huggingface.co/datasets/yatharth97/isa_gemma

User Guide

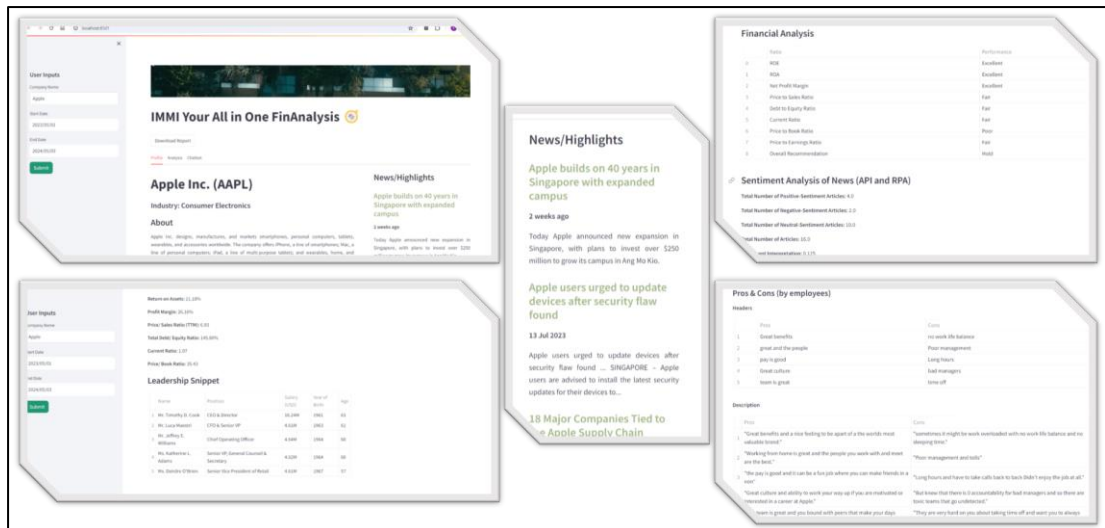
1. Launch the system following <Instructions on how to run Code> above.
2. Under User Inputs on the left of the webpage, key in the Company of interest and the respective start and end dates for required timeframe.



3. Click <Submit> button to launch the RPA/API process scrapping key company information, financial data and charts, followed by news, then top competitors, Glassdoor reviews, and finally ESG.
4. Once Scrapping is completed, webpage will reload to refresh the results.



5. Profile tab displays the Company Information, Financial Metrics including trend charts, key leadership personnel, as well as News highlights.



6. Analysis tab displays ESG Scores, Recommendation Ratings from key financial metrics, Sentiment Analysis from News, and Glassdoor employee reviews including qualitative feedback and comparison ranking against top competitors.

Financial Analysis			Performance
	Ratio		
0	ROE		Excellent
1	ROA		Excellent
2	Net Profit Margin		Excellent
3	Price to Sales Ratio		Fair
4	Debt to Equity Ratio		Fair
5	Current Ratio		Fair
6	Price to Book Ratio		Poor
7	Price to Earnings Ratio		Fair
8	Overall Recommendation		Hold

Sentiment Analysis of News (API and RPA)	
Total Number of Positive-Sentiment Articles:	4.0
Total Number of Negative-Sentiment Articles:	2.0
Total Number of Neutral-Sentiment Articles:	10.0

7. Report can also be downloaded by clicking on the <Download Report> button.
8. Chatbot can be used on the Chatbot tab to query directly on related questions.

REFERENCES

1. IBM. (2024). Top financial services trends of 2024.
<https://www.ibm.com/blog/financial-services-trends/>
2. PwC. (2024). Next in banking and capital markets 2024: Turn uncertainty into opportunity. <https://www.pwc.com/us/en/industries/financial-services/library/banking-industry-trends.html>
3. Business Partner Magazine. (2024). Key trends impacting the investment management industry.
https://issuu.com/businesspartnermagazine/docs/issue_56
4. Craft. (2024). S&P Global competitors and similar companies. Retrieved from <https://craft.co/s-p-global/competitors#>
5. Mordor Intelligence. (2023). Business analytics solutions market - Size, companies & trends. Retrieved from <https://www.mordorintelligence.com>
6. Grand View Research. (2021). AI in Fintech Market Size, Share & Trends Analysis Report
7. Allied Market Research. (2024). AI in fintech market size, share and analysis | Forecast - 2031. Retrieved from <https://www.alliedmarketresearch.com/ai-in-fintech-market>
8. McKinsey & Company. (2024). Impact of artificial intelligence in finance. Retrieved from <https://www.mckinsey.com/industries/financial-services/our-insights/impact-of-artificial-intelligence-in-finance>
9. Craft.co. (2024). Top S&P Global Competitors and Alternatives. [online] Available at: <https://craft.co> [Accessed 29 Apr. 2024].
10. Mordor Intelligence. (2023). Business Analytics Market - Growth, Trends, COVID-19 Impact, and Forecasts (2023 - 2028). [online] Available at: <https://www.mordorintelligence.com> [Accessed 29 Apr. 2024].