SLIIT UNI
THE KNOWLEDGE UNIVERSITY

IT4010 Research Project – 2026 Jan

Topic Assessment Form

V2.1

Project ID:

1. Topic (12 words max)

AI Powered Stock Market Insight Generator for the Colombo Stock Exchange

2. Research group the project belongs to

CoEAI - Centre of Excellence for AI

3. Specialization of the project belongs to

Information Technology (IT)

4. If a continuation of a previous project:

Project ID	
Year	

5. Brief description of the research problem including references (300 words max) – references not included in word count.

The Colombo Stock Exchange (CSE) has experienced a notable rise in participation from retail investors, particularly among younger demographics driven by digital accessibility and interest in alternative income streams. However, this growth has been accompanied by misinformation, emotionally driven decision making, and unverified advisory content that can mislead inexperienced investors. Many new investors lack financial literacy and struggle to interpret market data, annual reports, macroeconomic indicators, or geopolitical news. Without proper tools to understand risk, trends, and contextual signals, investors may make poor decisions resulting in financial loss, reduced market confidence, and increased vulnerability to fraudulent schemes.

Existing financial platforms primarily cater to advanced traders, offering technical charting tools without explainability or localized market context. Global stock analysis systems rarely account for emerging market characteristics such as political instability, currency fluctuations, regulatory changes, or country-specific financial reporting standards. Academic research similarly focuses heavily on predictive performance rather than interpretability or investor education. As a result, retail investors face a gap in obtaining transparent, data driven, and explainable insights tailored to Sri Lanka's market conditions.

This problem has economic and social implications. Reduced investor confidence negatively impacts liquidity and capital formation, while misinformation fuels speculative behavior. There is a need for an AI driven system that helps investors understand "why" market changes occur without functioning as a financial advisory service. Such a system can improve retail investor confidence, promote responsible participation, and support the development of a healthier capital market ecosystem.

This research aligns with SDG 8 (Decent Work and Economic Growth) by enabling informed financial engagement, and SDG 9 (Industry, Innovation and Infrastructure) by promoting responsible AI integration within a critical financial domain.

References

[1] Barber, B. M., & Odean, T. (2013). The behavior of individual investors. *Handbook of the Economics of Finance*, 2, 1533–1570.

[2] Shiller, R. J. (2017). Narrative economics. *American Economic Review*, 107(4), 967–1004.

[3] Bouteska, A., & Regaieg, O. (2020). Investor sentiment and stock market volatility. *Finance Research Letters*, 36, 101378.

[4] Molnar, C. (2022). *Interpretable Machine Learning*. Lulu Press. Available at https://christophm.github.io/interpretable-ml-book/

[5] Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

[6] Fama, E. F. (1998). Market efficiency long-term returns and behavioral finance. *Journal of Financial Economics*, 49(3), 283–306.

[7] World Bank. (2022). Financial consumer protection and fraud prevention in emerging markets. World Bank Publications.

[8] United Nations. (2015). Transforming our world: The 2030 Agenda for Sustainable Development. United Nations General Assembly.

6. Brief description of Existing Research and Systems (300 words max)

existing commercial platforms such as TradingView, MetaTrader, Yahoo Finance, and Investing.com provide technical indicators, charting tools, and basic screening features for global financial markets. While they are widely used, these platforms assume prior financial knowledge, do not provide explainable insights, and are not optimized for emerging market characteristics. Furthermore, they lack contextual interpretations based on macroeconomic events, geopolitical news, and financial document insights relevant to the Colombo Stock Exchange.

Academic research in stock market forecasting commonly explores machine learning techniques including ARIMA, LSTM, GRU, Random Forest, and Transformer architectures for price prediction and volatility estimation. Although these models achieve promising predictive performance, explainability is rarely emphasized. Investors may receive directional predictions without understanding underlying drivers or influencing factors. Explainable Artificial Intelligence (XAI) research such as SHAP and LIME has recently gained attention to address interpretability gaps, but has not been widely applied in emerging market finance.

In addition to quantitative price models, recent work in financial natural language processing explores sentiment extraction from news, policy statements, and earnings reports. However, these systems are predominantly developed for high-resource markets such as the US or EU exchanges. Large Language Models (LLMs) have shown strong capabilities in financial document understanding, yet domain adaptation and evaluation for developing markets remains limited.

No existing open source or academic system combines time-series ML models, sentiment/event signals, and LLM-based financial document understanding into a unified explainable insight pipeline for an emerging market such as Sri Lanka. This presents an opportunity to contribute both academically and practically by integrating multiple data modalities numerical, textual, and contextual into a transparent investor support system rather than a trading advisory tool.

7. Brief description of the solution's nature, including a conceptual diagram (maximum 500 words).

The proposed solution is an AI powered stock market insight generation system designed for the Colombo Stock Exchange. The system does not provide buy/sell recommendations, but instead produces interpretable insights that help retail investors understand market behavior, risk factors, and contextual influences.

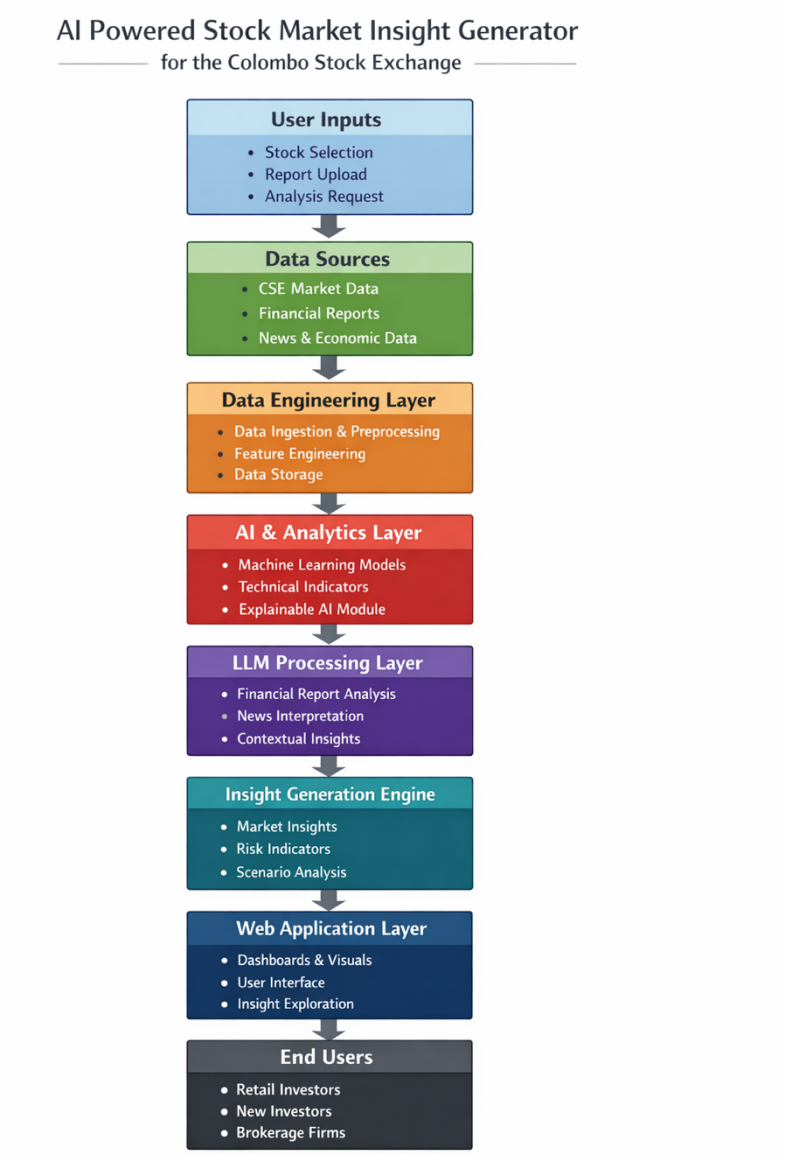
The system ingests three primary categories of data: (1) historical stock price and trading volume data, (2) financial documents such as annual reports and earnings statements, and (3) news articles and macroeconomic information. These data sources represent structured, semi-structured, and unstructured formats and are processed within separate analytical components.

A data engineering layer handles data collection, cleaning, transformation, and feature engineering. Time-series features, technical indicators, and financial ratios are generated to support modeling. Machine learning models analyze historical price and volume patterns to estimate trends, volatility, and risk. Explainable AI (XAI) methods such as SHAP are then applied to reveal the contribution of different features, enabling interpretability for non-technical users.

Natural language processing is used to extract sentiment and event signals from relevant news and policy updates. These signals help contextualize external drivers affecting stock movements. Additionally, a Large Language Model component processes financial reports to extract key metrics and generate simplified textual insights. This supports financial literacy by translating complex financial statements into understandable narratives.

The multi-modal insights from ML, sentiment analysis, and LLM processing are fused and presented through a web dashboard. Visualizations display trends, explanations, risk indicators, and contextual commentary. The user interface allows investors to explore results without requiring advanced financial knowledge. The system architecture supports local execution using MERN stack and Streamlit, enabling an accessible and cost-free development environment.

The conceptual diagram illustrates the end-to-end pipeline from user inputs and data sources through preprocessing, AI modeling, explainability, and final visualization layers.



8. Brief overview of the availability of necessary specialized domain expertise, knowledge, and data. (500 words max)

   Domain expertise will be acquired through academic literature, finance resources from the Colombo Stock Exchange, financial news outlets, and supervisor guidance. The team

includes an active retail investor, providing practical understanding of user pain points and decision-making behavior.

Public market data and company filings are readily available through the Colombo Stock Exchange website, PDF financial reports, and financial news services. NLP datasets for sentiment can be sourced through free APIs, web scraping, or open datasets. No sensitive personal data is involved and ethical clearance is not required.

The team possesses skills across data engineering, machine learning, NLP, LLM integration, frontend development, backend development, and system testing. The system can be developed entirely with open-source tools including MERN stack, Python, Streamlit, Hugging Face models, MongoDB Community Server, and News APIs.

9. Objectives and Novelty

Main Objective

To design and implement an AI powered stock market insight generation system that provides interpretable and data driven insights for the Colombo Stock Exchange.

Member Name with Registration No	Sub Objective	Tasks	Novelty
S A H Samarathunga (IT21296482)	Market Modeling and Explainable AI Insight Generation	• Design and implement end to end data pipelines for Colombo Stock Exchange market data • Collect preprocess and manage structured and unstructured financial datasets • Develop and train machine learning models for market trend analysis risk estimation and pattern detection • Apply explainable AI techniques to interpret and justify model outputs • Integrate large language models for financial report understanding and insight generation	End to end AI driven insight generation combining data engineering machine learning explainable AI and large language models tailored for a localized stock market context.

		• Expose AI model outputs through APIs to support backend and frontend integration	
K C S Fernando (IT21471162)	Financial Document Understanding and Language-Driven Insight Extraction	• Collect and preprocess company annual reports earnings statements and regulatory disclosures • Extract key financial indicators financial ratios and risk factors from PDF documents • Integrate large language models to summarize financial documents and generate simplified investor friendly insights • Benchmark performance of multiple LLM configurations for accuracy relevance and contextual correctness • Provide unstructured to structured financial insight outputs to support downstream analysis modules	Application of large language models for financial document understanding and structured insight generation in a localized emerging market financial environment.
T G R N S Wijesooriya (IT22547088)	Insight Visualization and Human Interpretability Evaluation	• Design and implement dashboards to visualize trends sentiment signals and explanations generated by AI modules	Human centered interpretability evaluation of AI generated financial insights enabling effective investor comprehension through visualization and usability driven refinement.

		• Integrate multi modal insights (ML + sentiment + LLM outputs) into a unified user interface • Conduct interpretability evaluation and user oriented testing to assess the clarity usefulness and cognitive load of generated insights • Refine interface based on feedback to enhance decision support for non expert retail investors • Support usability testing technical documentation and demonstration preparations	
Harithra K (IT22065308)	Sentiment and Event Driven Financial Insight Modeling	• Collect financial news policy updates and macroeconomic event data related to the Colombo Stock Exchange • Apply natural language processing techniques to extract sentiment and classify event impact signals • Analyze correlations between sentiment/event signals and subsequent stock price movements • Evaluate sentiment models and event driven features using statistical and financial performance metrics • Provide enriched contextual signals to complement machine learning model outputs	Introduction of sentiment and macro event driven financial insight modeling for an emerging market exchange enabling multi modal context aware interpretation of stock behavior.

SLIIT UNI
THE KNOWLEDGE UNIVERSITY

IT4010 Research Project – 2026 Jan
Topic Assessment Form
V2.1

10. This part is to be filled by the Supervisor and the Co-supervisor of the Project.

a) This research topic possesses a comprehensive scope suitable for a final-year project.

Yes	✓	No	

b) The proposed topic exhibits novelty.

Yes	✓	No	

c) The student group can successfully execute the proposed project.

Yes	✓	No	

d) The proposed sub-objectives reflect the students' areas of specialization.

Yes		No	✓

e) Any other comments:

	Title	First Name	Last Name	Signature
Supervisor	Dr.	Kapila	Dissanayaka	14/01/2026.
Co-Supervisor	Dr.	Mahima	Weerasinghe.	14/01/2026.
External Supervisor				
Summary of external supervisor's (if any) experience and expertise				

Page 10 of 10

V2.1 updated on 02nd Dec. 2025