

Vivekanand Education Society's Institute of Technology

Department of Computer Engineering

Group No. : 11

Date :- 02/08/2024

Project Synopsis (2024-25)

Financial Risk analysis using LLM

Dr. (Mrs) Sujata Khedkar,
Associate Professor,
Computer Engineering Department,
Vivekanand Education Society's Institute of Technology,
Chembur, Mumbai, 400074.

Srushti Satish Sambare
2021.srushti.sambare@ves.ac.in

Purtee Mahajan
2021.purtee.mahajan@ves.ac.in

Tasmiya Khan
2021.tasmiya.khan@ves.ac.in

Ketaki Nalawade
2021.ketaki.nalawade@ves.ac.in

Content

Abstract	2
Introduction	2
Problem Statement	3
Proposed Solution	4
Methodology / Block Diagram	6
Hardware, Software and tools Requirements	7
Proposed Evaluation Measures	8
Conclusion	9
References	9

1. Abstract

Earnings calls are hosted by management of public companies to discuss the company's financial performance with analysts and investors[1]. Earnings calls play a significant role in providing vital information to the stakeholders. This project enhances corporate call analysis by taking audio / youtube video link / transcript pdfs / news of companies in any duration and financial reports as input and generating transcript , transcript summarization, timeline analysis , data interpretation of charts, risk assessment of stocks with comparison between companies and investment prediction via argument mining [11]. By integrating these methods, valuable insights can be drawn from the earning calls, which will in turn, help the investors and other stakeholders in decision making.

2. Introduction

Earnings calls play an important role in boosting the investments in the company by conducting economic communication. Earnings conference calls follow the quarterly release of a firm's earnings, and have increased in popularity in recent years, mainly because of their ease of accessibility through modern communication mediums (e.g., applications like EarningsCast, interactive investor-relation websites)[3][4]. The purpose of these calls is to tell the market about the firm's future strategy and tactics, as well as to comment on the previous quarter's revenue streams and costs[3][4].

Technical and financial analysis of the company along with the fundamental analysis of the economy is to be taken into account while investing in the capital market[5]. A variety of factors influence investors' perceptions of various investing options[5]. Data visualisation and interpretation are becoming increasingly crucial in today's commercial world [6]. Risk assessment involves identifying, analysing, evaluating potential risks that could negatively impact investment outcomes ,stock predictions and comparison between the companies using the news of companies in different durations[13] [11]. Therefore, developing an analyzer for these earning calls has become crucial in order to help the investors to make profitable decisions. Investment prediction aims to forecast future market trends and asset prices to inform investment strategies. By using FinCalls , investors can gain deeper insights into the factors driving market trends, thereby enhancing the accuracy of their predictions. It becomes easier for investors to analyse the financial growth of the companies using FinCalls.

3. Problem Statement

It has become crucial to analyse multiple data types such as quantitative data like stock prices and financial indicators and qualitative data such as financial news articles, earnings calls or transcripts, into a unified analysis for financial risk prediction.

Earnings calls last for a long duration. Understanding and remembering all the crucial information becomes a tedious job. This causes information overloading. Some information might be overwhelming especially for an investor who is new to the company or the industry.

Financial reports of a company, such as annual reports, quarterly reports, and other regulatory filings, provide information regarding a company's financial health, performance, and strategic initiatives. It takes a lot of time to study such a large report and find key information from it. It is observed that investors are more devoted and fond of finicky type of investment choice and preferences[5]. Forecasting stock trends is particularly significant, as it leverages historical data to shape trading strategies and pinpoint opportunities for buying or selling stocks [15] [13].

[13] Financial news plays a crucial role in decision-making processes across the financial sector, yet the efficient processing of this information into a structured format remains challenging. Timely and accurate analysis of financial news can provide valuable insights for market participants and researchers alike. However, extracting actionable information from the vast amount of unstructured news data remains a significant challenge [12][14].

The 'disclosure bias' is another major issue in earning calls. Some businesses may emphasise just the positive aspects while downplaying the unfavourable information. As a result, investors may struggle to gain a clear image of the company. These changes can be costly to investors, resulting in greater trading fees, missed purchasing opportunities, or overall position losses [1]. During the earnings calls, some analysts often ask questions that would promote their own investment opinions. This may influence the investors and also give biased information regarding the company.

[13] Risk assessment is crucial for investors, analysts, and companies alike. By analyzing the language used, tone, and content of earning calls, financial report data and real time financial news we can identify potential risks that could impact the company's financial health and stock price. By analyzing language patterns and argument structures, it's possible to detect early warning signs, understand risk over different time frames, operational issues, or legal challenges.

4. Proposed Solution

The proposed system - 'Fincalls - Risk Analyzer' acts as an effective solution for the economic development of the company as well as provide an ease of analysis to the investors.

Following sources of data are taken into consideration.

1. Corporate Earnings Calls Audio

- a. The corporate earnings call audio is available on the company websites. It can be uploaded. The tone and semantic analysis of the call will be done and a risk analysis will be given.
- b. Also, a transcript of the call will be generated.
- c. Using this generated transcript, the summary of the call as well as the timeline analysis of the call will be given.

2. Corporate Earnings Calls Youtube Video URL

- a. Along with the youtube channels owned by the companies, there are multiple channels that post earnings calls video recordings on youtube.
- b. Again, after extracting the audio, the tone and semantic analysis of the call will be done and a risk analysis will be given.
- c. Using this generated transcript, the summary of the call as well as the timeline analysis of the call will be given.

3. Corporate Earnings Calls Transcript PDF

- a. Sometimes, even the direct transcript is available instead of the call.
- b. From this transcript, the summary of the call as well as the timeline analysis of the call will be given.
- c. Also, via argument mining, the risk analysis will be given.

4. Company News

- a. The company name and the duration of the release of the news articles will be given.
- b. And from this, the relevant news will be extracted.
- c. A risk analysis will be given corresponding to the extracted news.

5. Financial Report

- a. A search bar will be provided, where, after giving the company name and the year as an input, the financial reports of the company will be fetched.
- b. They can be downloaded and used further.
- c. Once uploaded, a risk analysis of the reports will be provided, too.

Note that, the uploaded data should belong to surrounding durations. That is, the quarter to which the earnings call belongs should be a part of the year to which the financial report belongs. Also, the news duration should lie in the same duration.

Now, the risk analysis from all the above mentioned sources will be combined to form a Risk Report. This report can be downloaded in a PDF format.

The Risk Report will be created based on the following aspects:

1. Financial Performance Metrics: To analyze metrics focusing on revenue growth, profitability, cash flow, and debt levels and identify any potential risks associated.
2. Corporate Strategy and Governance: Review recent changes in leadership or corporate structure to identify potential risks associated with these changes.
3. Capital Allocation and Dividends: It looks at the company's strategies for reinvesting profits, paying dividends, conducting share buybacks, and managing debt.
4. Product and Market Focus: Evaluate how well the company's products and market strategy align with current and future market trends.
5. ESG and Social Impact: Examine the company's ESG reports, CSR initiatives, and sustainability goals to identify any gaps or areas where the company is underperforming.
6. Forward-Looking Statements and Guidance: Identify the company's future expectations by reviewing management's guidance on growth prospects, potential risks, and strategic priorities highlighted during the earnings call.

Another part is that the interpretation of dynamically generated graphs will also be given.

5. Methodology / Block Diagram

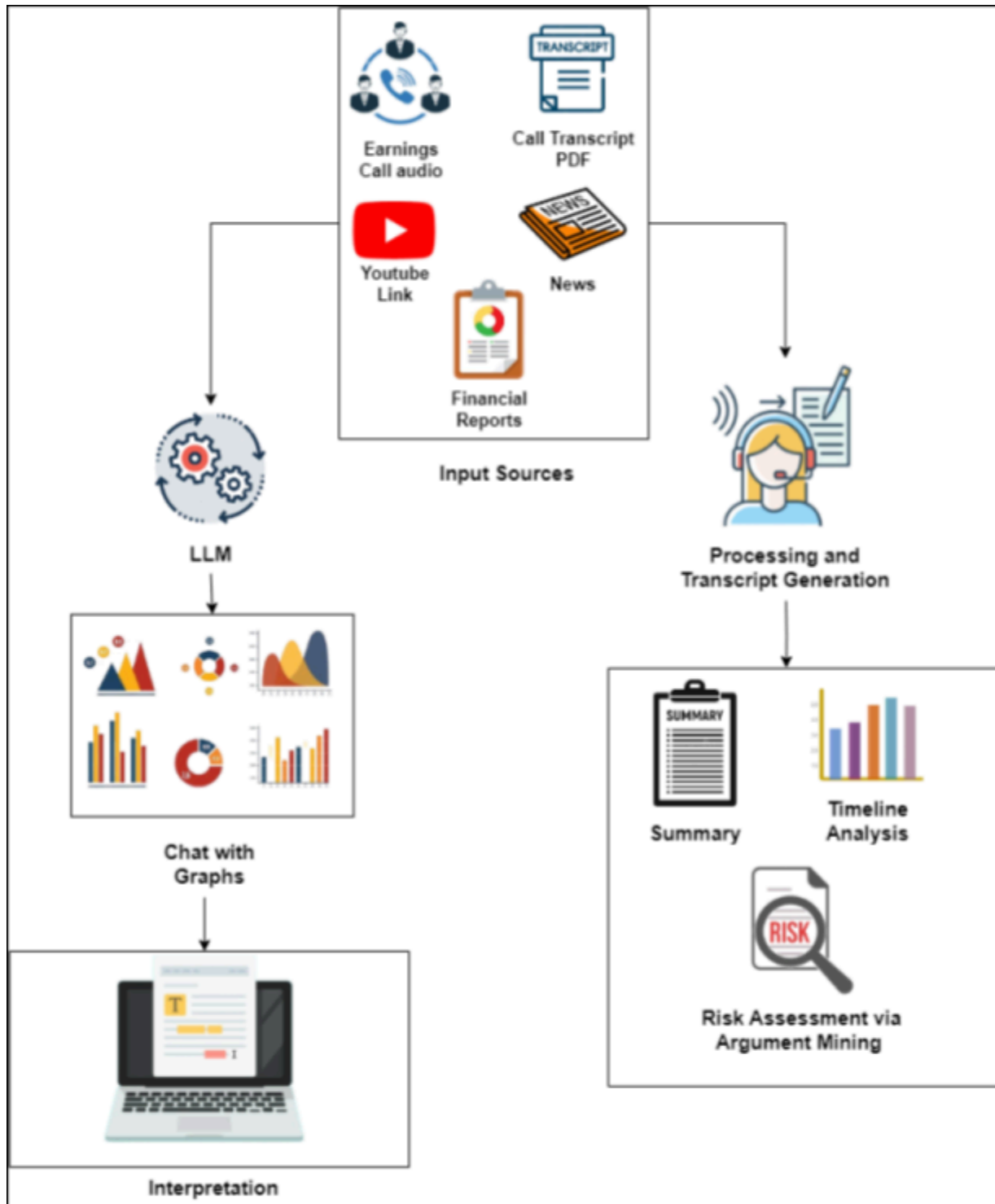


Fig.2 Block Diagram of the System

The user will have to upload either or all of the supported input sources which comprise of the Earnings Call audio, Earnings Call Transcript, Earnings Call YouTube Video URL, News and financial reports. Then, the user can either go for ‘Chat with Graphs’ or ‘Risk Analysis’ module.

If the user chooses the former, then, the numerical and the statistical data would be extracted from the given sources and would be represented in a graphical format. The interpretation of the formed graphs will be given and the user can interact and chat with those visualisations.

However, if the user chooses the latter, then, a separate risk analysis will be done on each source and the combined results will be used to form a risk report. Along with that, some other features include summary and timeline analysis of the Earnings Call.

6. Hardware, Software and tools Requirements

1. Frontend:

The frontend will be done using React.js - a JavaScript Frontend Framework which will provide a user interface for our project.

2. Transcribing the source earning call:

Google Text-to-Speech: We will use this ASR (Automatic Speech Recognition) API to transcribe the uploaded source file.

3. Summarization of transcript:

We will use argument mining to extract and summarize key arguments from the transcribed text. This method identifies core arguments and supporting evidence. Advanced models like BERT or Gemini will be employed to enhance the summarization, with BERT providing bidirectional context and Gemini offering deeper argument analysis. This approach ensures a focused and insightful summary of complex discussions.

4. Extracting Financial Data:

NLP Model and Named Entity Recognition (NER): For the extraction of financial data from our transcripts, we will require NER techniques. The pre-trained model can be used which has NER techniques, but it should be trained specifically for the extraction of financial data. We need to train the model for extracting quantitative measures, keywords related to revenue, earnings per share, revenue generated by the company and a lot more related parameters. This model will identify all the texts that match financial measures and parameters from the generated transcripts.

5. **Data interpretation of charts :** Users can chat with charts or graphs, which the Gemini model will analyze to extract meaningful insights and trends, enabling a deeper understanding of complex data and supporting informed decision-making.
6. **Risk assessment and Investment prediction:** Using either the Gemini or PaLM model, we will analyze argumentative structures in financial reports, news to assess risks and predict investment outcomes, helping users make more informed decisions based on market insights.
7. **News as input :** Integrate media news using a news API to evaluate the current economic, political, and social climate impacting market sentiment and investor behaviour. This helps gauge how events and trends influence market dynamics and investment decisions.
8. **Financial Report Retrieval and Risk Analysis:** A search bar will be available for users to input a company name and year, allowing them to fetch and download the company's financial reports using the Google Custom Search API. Once the reports are uploaded, the system will provide a risk analysis, offering insights into potential financial risks. This feature streamlines the process of accessing and evaluating financial data, supporting informed decision-making and thorough risk assessment.

7. Proposed Evaluation Measures

a. Transcript Accuracy:

To make sure that the generated transcript is accurate. The following factors that measures the transcript accuracy are:

- i. **Audio quality:** The audio quality of the source file will affect the accuracy of the generated transcript. The audio should be clear and of high-quality. The low-quality and noisy audio will significantly affect the accuracy of the transcript.
- ii. **Speaker Variation:** When there are multiple speakers speaking simultaneously. This will be difficult for the ASR system to recognize the voices.

b. Summarization Accuracy:

- i. **Model:** The proposed system will use a pre-trained model like BART which generates high-quality summaries.
- ii. **Training the Model:** The pre-trained model will fine-tune on the financial related phrases and terminologies which will help to generate more accurate summaries.

- iii. **Rouge Score:** Rouge score is used to evaluate the quality of the NLP tasks, here, transcript summarisation. The human generated summaries are known as references, while those generated by the model are known as candidates. Both recall and precision are used to compare the candidates and references.

c. Evaluation measures for NER:

- i. **Precision:** Precision will measure the accuracy of the NER system to identify the entities correctly. It is the ratio of correctly identified entities to the total number of entities predicted by model.
$$\text{Precision} = (\text{True Positives}) / (\text{True Positives} + \text{False Positive})$$

True Positives: The number of entities correctly identified by NER.
False Positives: The number of entities incorrectly identified by NER.
- ii. **Recall:** The recall metric reveals the number of predicted entities that are correct. It is the ratio of the number of correctly identified entities to the total number of entities present in the text.
- iii. **F1-score:** The F1 score is the function of precision and recall. It is used to balance the trade-off between precision and recall.
$$\text{F1-score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}). [9][10].$$
- iv. **Risk Prediction Accuracy:** Evaluate how accurately the extracted arguments predict actual risk events by comparing model predictions with historical data.

d. Argument Mining for Investment Prediction:

- i. **Sentiment Correlation:** Measure the correlation between extracted argumentative structures and market sentiment indicators to evaluate predictive power.

8. Conclusion

We aim to develop a 'Fincalls - Earnings Calls Analyzer' website which will take audio / youtube video link / transcript pdfs / news of companies in any duration and financial reports as input and generate transcript , transcript summarization, timeline analysis , data interpretation of charts, risk assessment of stocks with comparison between companies and investment prediction via argument mining.

9. References

- [1] Zhiqiang Ma, Grace Bang, Chong Wang & Xiaomo Liu. (2020), "Towards Earnings Call and Stock Price Movement"
- [2] Narayanan Bhaskaran, Dr.S.Suriyamurthi & M.Arivazhagan. (2012), "Investors Behaviour in Various Investment Avenues A Study", *Article in International Journal of Technology Marketing* · July 2012
- [3] Vivien E. Jancenelle Susan Storrud-barnes Rajshekhar Javalgi , (2017)," Corporate Entrepreneurship and Market Performance: A Content Analysis of Earnings Conference Calls ", *Management Research Review*, Vol. 40 Iss 3 pp. -
- [4] Kimbrough, M.D. (2005), "The effect of conference calls on analyst and market underreaction to earnings announcements", *The Accounting Review*, Vol. 80 No. 1, pp. 189-219.
- [5] Jeet Singh & Preeti Yadav. (2016), "A Study on the Factors Influencing Investors Decision in Investing in Equity Shares in Jaipur and Moradabad with Special Reference to Gender", *Amity Journal of Finance 1(1)*, (117-130) ©2016 ADMAA
- [6] Michael Diamond & Angela Mattia. "Data visualization: an exploratory study into the software tools used by businesses", *Journal of Instructional Pedagogies Volume 18*
- [7] Matthew N. O. Sadiku, Adebawale E. Shadare, Sarhan M. Musa & Cajetan M. Akujuobi. (2016), "Data Visualization", *International Journal of Engineering Research And Advanced Technology(IJERAT) ISSN: 2454-6135 [Volume. 02 Issue.12, December– 2016]*
- [8] Grishman, Ralph; Sundheim, B. 1996. Message Understanding Conference-6: A Brief History. *In Proc. International Conference on Computational Linguistics*.
- [9] Chinchor, Nancy. 1999. Overview of MUC-7/MET-2. *In Proc. Message Understanding Conference MUC-7*.
- [10] Nadeau, David; Sekine, Satoshi (2007). "A survey of named entity recognition and classification", *Lingvisticae Investigationes ISSN:0378-4169 [Volume 30, Issue 1, Jan 2007, p. 3 - 26]*
- [11] John Lawrence, Chris Reed; Argument Mining: *A Survey. Computational Linguistics* 2019; 45 (4): 765–818.
- [12] Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. 2015. Deep learning for event-driven stock prediction. In Twenty-fourth international joint conference on artificial intelligence
- [13] Cao Y, Chen Z, Pei Q, Dimino F, Ausiello L, Kumar P, Subbalakshmi KP, Ndiaye PM. RiskLabs: Predicting Financial Risk Using Large Language Model Based on Multi-Sources Data. arXiv preprint arXiv:2404.07452. 2024 Apr 11.
- [14] Dolphin R, Dursun J, Chow J, Blankenship J, Adams K, Pike Q. Extracting Structured Insights from Financial News: An Augmented LLM Driven Approach. arXiv preprint arXiv:2407.15788. 2024 Jul 22.

[15] Wang M, Izumi K, Sakaji H. LLMFactor: Extracting Profitable Factors through Prompts for Explainable Stock Movement Prediction. arXiv preprint arXiv:2406.10811. 2024 Jun 16.