InFiNITE (∞): Indian Financial Narrative Inference Tasks & Evaluations

Anonymous ACL submission

Abstract

This paper introduces **In**dian **Fi**nancial **N**arrative **I**nference **T**asks and **E**valuations (**InFiNITE**), a comprehensive framework for analyzing India's financial narratives through three novel inference tasks. Firstly, we present multi-modal earnings call analysis by integrating transcripts, presentation visuals, and market indicators via the **Mu**lti-**M**odal **I**ndian Earnings **C**alls (**MiMIC**) dataset, enabling holistic prediction of post-call stock movements. Secondly, our **B**udget-**A**ssisted **S**ectoral **I**mpact **R**anking (**BASIR**) dataset aids in systematically decoding government fiscal narratives by classifying budget excerpts into 81 economic sectors and evaluating their post-announcement equity performance. Thirdly, we introduce **B**harat **I**PO **R**ating (**BIR**) datasets to redefine Initial Public Offering (IPO) evaluation through prospectus analysis, classifying potential investments into four recommendation categories (Apply, May Apply, Neutral, Avoid). By unifying textual, visual, and quantitative modalities across corporate, governmental, and public investment domains, **InFiNITE** addresses critical gaps in Indian financial narrative analysis. The open source data sets of the framework, including earnings calls, union budgets, and IPO prospectuses, establish benchmark resources specific to India for computational economic research under permissive licenses. For investors, **InFiNITE** enables data-driven identification of capital allocation opportunities and IPO risks, while policymakers gain structured insights to assess Indian fiscal communication impacts. By releasing these datasets publicly, we aim to facilitate research in computational economics and financial text analysis, particularly for the Indian market.

1 Introduction

In financial markets, comprehensive analysis of diverse narratives forms the foundation of informed decision-making. Whether in corporate earnings presentations, government budget announcements, or IPO prospectuses, extracting actionable insights from complex financial narratives remains a significant challenge in the Indian context. This paper introduces **InFiNITE** (**In**dian **Fi**nancial **N**arrative **I**nference **T**asks and **E**valuations) (Figure: 1), a framework addressing three critical domains of financial narrative analysis through specialized computational approaches.

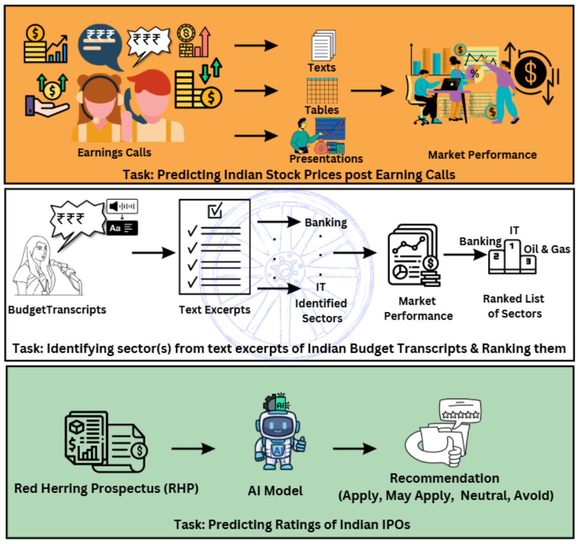

Figure 1: Indian Financial Narratives Analytics Tasks

Financial narrative analysis in India faces three key challenges: 1) Corporate earnings calls integrate multiple data types, but research lacks multimodal approaches that combine text, visuals, and tables, especially for Indian markets; 2) Union Budget analysis remains manual despite significantly impacting sectoral performance and market volatility[1], with investors struggling to process complex fiscal implications[2]; and 3) IPO prospectuses (80-

[1] https://cleartax.in/s/budget-day-market-movement-history-in-india

[2] https://economictimes.com/markets/stocks/news/consumption-over-capex-how-the-budget-impacts-stock-market-investors/articleshow/117853360.cms

300 pages) overwhelm individual investors, particularly since Securities and Exchange Board of India (SEBI) made professional grading optional in 2014[3]. These domains urgently need automated, objective analytical tools.

The **InFiNITE** framework addresses these challenges by developing specialized computational approaches for each financial narrative domain.

Our Contributions

Our primary contributions include:

- **MiMIC Dataset**: We introduce the first multi-modal dataset (**M**ulti-**M**odal **I**ndian Earnings **C**alls) comprising earnings call transcripts, presentations, fundamentals, technical indicators, and post-announcement stock price data from Indian companies.

- **BASIR Dataset**: We present **B**udget-**A**ssisted **S**ectoral **I**mpact **R**anking, the first annotated dataset spanning Indian Union Budgets from 1947 to 2025, featuring 1,600+ labeled budget transcript excerpts and 400+ texts with corresponding post-budget sectoral performance metrics.

- **BIR Datasets**: We introduce two India-specific datasets (**B**harat **IPO** **R**ating) for Main Board (MB) and Small and Medium Enterprises (SME) IPOs, enabling automated prospectus analysis and investment recommendation.

- **Integrated Analytical Frameworks**: We develop specialized computational approaches for each domain: (1) a multi-modal framework for earnings call analysis, (2) a sector identification and ranking system for budget analysis, and (3) a Retrieval Augmented Generation (RAG) framework for IPO prospectus mining that outperforms state-of-the-art Large Language Models.

Through these contributions, **InFiNITE** establishes benchmark resources for computational economics research while providing practical insights to decode India's complex financial narratives, enhancing decision-making capabilities for investors, analysts, and policymakers.

[3] https://www.angelone.in/knowledge-center/ipo/ipo-grading

2 Related Work

2.1 Analysis of Corporate Earnings Calls

The analysis of earnings calls for stock price prediction has evolved into a prominent research area, driven by advancements in multi-modal data integration. Earnings calls serve as vital information repositories, offering insights beyond conventional financial indicators. Research by Medya et al. (Medya et al., 2022) demonstrates the predictive power of semantic elements within earnings call transcripts, showing that narrative structure and tonal qualities of these corporate communications substantially shape investor sentiment and subsequent market reactions. Complementing this, Huynh and Shenai (Huynh and Shenai, 2019) document an inverse relationship between option trading volumes and immediate stock price reactions following earnings announcements.

Early approaches to earnings call analysis relied on textual sentiment analysis using financial-specific dictionaries (Loughran and McDonald, 2011). A significant breakthrough came with models jointly analyzing verbal and vocal cues. Qin and Yang (Qin and Yang, 2019) proposed a deep learning framework combining textual content with acoustic features, demonstrating that how executives speak significantly impacts market response. Building on this foundation, Sawhney et al. (Sawhney et al., 2020a) introduced a neural architecture employing cross-modal attention mechanisms to capture verbal-vocal coherence while incorporating stock network correlations through graph-based learning.

Research has further evolved to include vocal/audio analysis of manager speech patterns (Sawhney et al., 2021a), Graph Neural Networks for text classification, and combined verbal-vocal cue analysis for volatility (Sawhney et al., 2020b) and risk prediction (Sawhney et al., 2020a). However, these approaches have predominantly focused on US markets, with limited research specifically addressing Indian earnings calls. The distinct characteristics of Indian financial markets—including regulatory variations, cultural nuances in communication, and unique market dynamics—necessitate tailored approaches rather than direct adoption of models designed for Western markets.

2.2 Impact of Budget on Financial Markets

The annual Indian Union Budget functions as a crucial economic policymaking instrument, directly impacting sectoral growth trajectories and investor sentiment in equity markets (Panwar and Nidugala, 2019). Event studies have demonstrated that Cumulative Average Abnormal Returns (CAARs) are significant around budget announcements, indicating that these events contain valuable information for market participants (Kharuri et al., 2021; Manjunatha and Kharuri, 2023).

Studies by Martin et al. (Martin, 2024) and Joshi et al. (Joshi and Mehta, 2018) reveal pronounced sector-specific volatility patterns post-budget announcements, with healthcare, banking, and Information Technology sectors demonstrating heightened sensitivity to tax reforms and capital allocation decisions. This sector-specific analysis is particularly relevant to the Indian stock market, where finance and services sectors frequently dominate overall market performance.

Natural Language Processing (NLP) has emerged as a transformative tool in decoding fiscal policy impacts on stock markets. Mansurali et al. (Mansurali et al., 2022) analyzed sentiments of tweets relating to Budget 2020, while sentiment analysis has proven useful in assessing market sentiment and generating trading signals based on prevailing trends (Saxena et al., 2021). Advanced NLP models like BERTopic (Grootendorst, 2022) and RoBERTa (Liu et al., 2019) have been employed to analyze the Reserve Bank of India's monetary policy communications, revealing how different economic topics influence market reactions (Kumar et al., 2024).

Most previous studies have focused on post-hoc analyses using historical data. Our work introduces a predictive approach, utilizing NLP to automatically detect sectors from budget announcements and rank them according to predicted performance, enabling proactive identification of potential market impacts.

2.3 IPO Rating Prediction

The prediction of Initial Public Offering performance has garnered significant attention, particularly due to its implications for investors and market efficiency. Most prior studies have concentrated on short-run underpricing (Anand and Singh, 2019; Bajo and Raimondo, 2017) or long-run underperformance (Sahoo and Rajib, 2010).

Several researchers have explored IPO grading's usefulness. Sarin (Sarin and Sidana, 2017) indicates that many retail investors are familiar with the IPO grading process, though perceptions of its effectiveness vary. While IPO grading is considered valuable for investors (Deb and Marisetty, 2010), its impact is inconsistent across different investor segments. Poudyal et al. (Poudyal, 2008) observed that securities with higher IPO grades exhibit lower degrees of underpricing and increased subscription rates across all investor types.

The influence of credit ratings on IPO underpricing has been well-documented. Dhamija and Arora (Dhamija and Arora, 2017) found that firms with credit ratings experience significantly less underpricing than those without, indicating that improved corporate governance and transparency can lead to better IPO valuations. Jacob and Agarwalla (Jacob and Agarwalla, 2015) explored mandatory IPO grading effects in India, concluding that such certifications can enhance institutional investor demand, though their impact on overall pricing efficiency is limited.

While these studies highlight IPO grading's significance, none propose automated methods for grading IPOs. Automated methods for predicting ratings from texts (Khan et al., 2021) have been well-studied in domains like e-commerce (Qu et al., 2010) and local services (Lei et al., 2016), but their application to IPO prospectuses represents a novel contribution to this field.

Our work addresses this gap by introducing a task for predicting ratings based on the prospectuses of Indian companies preparing for IPOs, presenting valuable insights that empowers investors with data-driven insights for more informed IPO subscription decisions.

3 Tasks and Datasets

3.1 Stock Price Prediction from Multi-Modal Earnings Calls (MiMIC)

3.1.1 Task

This study addresses the problem of predicting opening stock prices for Indian companies on the day following the release of quarterly earnings results, leveraging multi-modal data (numeric, text transcripts, images from presentations, and tabular data). The performance of the proposed framework is evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).

3.1.2 Dataset Construction

The **MiMIC** (**M**ulti-**M**odal **I**ndian Earnings **C**alls) dataset was constructed by systematically collecting and processing multi-modal data from earnings calls of Indian companies across different market capitalizations. This comprehensive dataset includes earnings call transcripts, presentation materials, fundamentals, technical indicators, and stock performance metrics to facilitate the analysis of market reactions following corporate disclosures.

Company Selection

We selected all companies representing the Nifty 50 Index, Nifty Midcap 50 index, and Nifty Smallcap 50 index of the Indian stock market as of 3rd November, 2024. For each company, we collected their NSE ticker symbols from their respective company profile pages, which served as unique identifiers throughout our data collection process. We had to eliminate certain companies due to the non-availability of sufficient information. Finally, we were left with 133 companies.

Multi-Modal Data Collection

For each selected company, we gathered the following data components from January 2019 to November 2024:

- **Textual Data:** Earnings call transcripts were collected from Screener.in [4] Text-heavy slides underwent Optical Character Recognition (OCR) to extract textual information.
- **Visual Data:** Presentation slides used during earnings calls were collected from the same website and visual elements such as charts, graphs, and images were preserved in their original format for visual analysis.
- **Tabular Data:** Financial tables from presentations were extracted separately using image2table [5] to maintain their structural integrity, as they often contain critical quantitative information about company performance.
- **Numeric Data:** We incorporated a range of numerical features, encompassing technical and fundamental indicators, macro-economic variables and market data, into our analysis. A comprehensive set of these variables as presented in §C.2

[4] https://www.screener.in/ (accessed on 30th November, 2024)

[5] https://github.com/xavctn/img2table (accessed on 28th March, 2025)

Stock Performance Data

To establish the relationship between earnings calls and subsequent market reactions, we collected stock price data for each company:

- Opening price on the day of earnings call (d)
- Opening price on the day following [6] earnings call ($d+1$)

We attempted to collect audio data for earnings calls, but it was unavailable in the majority of cases. The initial dataset underwent a cleaning process to remove instances where both the earnings call transcript and the corresponding presentation slides were unavailable. This resulted in a final dataset of 1,042 instances, derived from 768 transcripts and 833 presentations.

To evaluate the performance of the proposed models, we partitioned the dataset into three distinct subsets based on temporal criteria. Data spanning up to February 7, 2024, was allocated to the training set (80% of the total data). Data from February 8, 2024, to August 9, 2024, was used for validation (10%), and data beyond August 10, 2024, was reserved for testing (10%).

3.2 Sector Identification & Performance Prediction from Budgets (BASIR)

3.2.1 Tasks

This study addresses two sequential challenges in computational fiscal analysis:

1. **Multi-Label Sector Classification**
   Given excerpts from a budget transcript $t \in T$ from India's Union Budget corpus (1947–2025), determine the probabilistic association $P(s_i|t)$ for each sector $s_i \in S$, where $S = \{s_1, \ldots, s_{81}\}$ represents formal economic sectors. The task requires overcoming:
   - Implicit sector references in policy language (e.g., "Credit access for handloom industries" $\rightarrow$ Banking, Textile sectors)
   - Domain-specific lexical ambiguity (e.g., "digital infrastructure" mapping to both Technology & Utilities sectors)

[6] Note: We are using opening price of the next day and not the closing price of the day of earnings call because most of these calls happen after the market hours https://www.etnownews.com/markets/tcs-infosys-wipro-hcl-tech-q4-results-fy-2025-date-time-dividend-update-quarterly-earnings-schedule-article-151356517

2. **Performance-Aware Sector Ranking** For identified sector set $\hat{S} = \{s_j \mid P(s_j|t) > \tau\}$, develop a model $f : \hat{S} \rightarrow \mathbb{R}^+$ that ranks sectors by expected next day post-announcement returns r_s using text excerpts t related to the sector s_j. Here, τ represents probabilistic threshold.

We used F1 (Micro, Macro, Weighted) and Normalized Discounted Cumulative Gain (NDCG) scores for evaluating the classification and ranking problems respectively.

3.2.2 Dataset Construction

- **Sector-Company Mapping**: We systematically collected a list of sectors and their constituent companies from Screener.in. [7]
- **Budget Transcripts**: Aggregated 97 Union Budget documents (1947–2025) from India's Ministry of Finance portal [8], comprising 1,600+ text excerpts. This also includes the interim budgets.

Annotation Pipeline

1. **Sector Tagging**: For each of the budget transcripts, we prompted DeepSeek (DeepSeek-AI, 2025) to extract texts and corresponding sector(s) as mentioned in §D.3.1.
2. **Validation**: We manually validated all the outputs.

Market Response Quantification

For sector s in budget day d of a financial year, performance metric $r_{s,d}$ calculated as:

$$r_{s,d} = \frac{1}{|C_s|} \sum_{c \in C_s} \frac{P^{\text{open}}_{c,d+1} - P^{\text{open}}_{c,d}}{P^{\text{open}}_{c,d}}$$

where C_s denotes constituent companies of sectors, with historical data sourced from yahoo finance. [9] $P^{\text{open}}_{c,d}$ denotes the opening price of company c on day d. Finally, we ranked the sectors in decreasing order of their performances. More details about the data is presented in Table 2. Data till the year 2019 was used for training, data spanning 2020 to 2023 was allocated for validation, and 2024 data was reserved for testing.

[7] https://www.screener.in/explore/ (accessed on 17th March, 2025)

[8] https://www.indiabudget.gov.in/bspeech.php (accessed on 17th March, 2025)

[9] https://finance.yahoo.com/ (accessed on 17th March, 2025)

3.3 IPO Rating Prediction from Red Herring Prospectus (BIR)

3.3.1 Task

Given a company's IPO prospectus, our objective is to comprehend its content and categorize it into one of four classifications: Apply, May Apply, Neutral, or Avoid, providing a concise and informed assessment of the investment opportunity. As this is a classification problem with class imbalances, we used Micro, Macro, and weighted F1 scores for evaluation.

3.3.2 Dataset Construction

We introduce two new datasets for this task: one for MB IPOs and another for SME IPOs, each serving distinct market segments. Mainboard IPOs are intended for larger, established companies, while SME IPOs cater to smaller enterprises. We gathered data on MB and SME IPOs separately from the chittorgarh website.[10] The MB data is available from 2011, while SME data starts from 2012. Our collection of this data continued until November 7, 2024, and includes the following information: Review Title (this contains the name of the company as well), Year of the IPO, Link to access the review, Link to a webpage containing comprehensive details about the IPO, Key (Unique identifier of each row), Link to access the (D)RHP in PDF format, Name of the JSON file having text contents extracted from (D)RHP, Text content of the review, Recommendation (Apply, May Apply, Neutral, or Avoid). We removed the author names to maintain annonimity.

We excluded entries without reviews or recommendations. Notably, MB IPOs often have multiple reviews; in such cases, we retained only those reviews that aligned with the majority recommendation. For example, if a company has five reviews—three recommending "Apply" and two recommending "Avoid"—we would keep only the three "Apply" reviews. Conversely, 97% of SME IPOs have reviews authored by a single individual, leading us to discard the remaining 3% of data. For reviews provided in PDF format, we utilized PyPDF [11] to extract text. The Draft Red Herring Prospectuses (DRHP) and Red Herring Prospectuses (RHP), were available in PDF format. In instances where both DRHP and RHP were present,

[10] https://chittorgarh.com/ (accessed on 19th January, 2025)

[11] https://pypi.org/project/pypdf/ (accessed on 19th January, 2025)

we prioritized the RHP. To ensure the quality of our data, we compared IPO ratings with their actual opening prices. For Main Board IPOs, we found that in 82.17% of cases, an 'Apply' recommendation corresponded to an opening price higher than the issue price. For SME IPOs, it was 83.49%. In total, we collected 1,830 instances for mainboard IPOs and 1,131 for SME IPOs. Data up to 2023 was used for training purposes, while data from 2024 was reserved for testing.

4 Experiments and Results

4.1 Stock Price Prediction from Multi-Modal Earnings Calls (MiMIC)

Our experimental approach progressed through the following stages of feature incorporation:

1. **Numeric Features**: We initially utilized only numeric features (N). We trained various machine learning models (like Extreme Random Forest (Geurts et al., 2006), Distributed Random Forest (DRF) (H2O.ai, 2025), XGBoost (Chen and Guestrin, 2016), Gradient Boosting Machine (Friedman, 2001), Feed forward neural network based Deep Learning (**DL-1**), etc.) for regression using the AutoML framework of H2O. [12] The **DL-1** model performed the best.

2. **Text Features**: We expanded our feature set by incorporating textual data (T) from transcripts, presentations, and tables in markdown format. To represent these textual features, we employed the Nomic 1.5 (Nussbaum et al., 2024) model to extract embeddings (Em). We used matryoshka representation learning to truncate the dimension of embeddings to 128. This was essential as we had only 832 instances to train the regression models. After evaluating multiple H2O AutoML models, the feed-forward neural network (**DL-2**) demonstrated superior performance. Subsequently, we trained a XGBoost model for binary classification utilizing exclusively text embedding features to predict whether the stock's opening price on day (d+1) would exceed that of day (d). Its F1 score on validation set was 0.675. The predicted probability (P) outputs from this classifier were then incorporated as features in the original regression framework (**DL-1**), thereby creating a cascaded prediction framework. After, training multiple models using H2O AutoML, we obtain best results from a feed forward neural network based model (**DL-3**).

3. **Image Features**: We further augmented our dataset with visual information (I). We used the Nomic Vision 1.5 model (Nussbaum et al., 2024) to extract embeddings from images. For instances with multiple images, we applied mean pooling to the image embeddings. Just like the text embeddings, we truncated the dimension of embeddings to 128. Among H2O AutoML models trained on numeric data along with text and image embeddings taken together, the feed-forward neural network (**DL-4**) yielded optimal results. Following our text-based approach, we similarly trained a DRF model for binary classification using only image embeddings to predict next-day price increases. The F1 score of this classifer was 0.680. The resulting probability estimates were then used as features, in our regression framework (**DL-3**), extending our cascaded framework from numeric and text to visual data. We followed an identical evaluation process using H2O AutoML, with a feed-forward neural network (**DL-5**) similarly emerging as the optimal model, mirroring our findings from the text modality.

This stepwise approach allowed us to assess the impact of each feature type on the model's performance. Finally, we evaluated the performance of Llama-4 Maverick (Meta AI, 2025), a state-of-the-art multi-modal vision language model, under zero-shot conditions (§C.5) using raw images and text. The results corresponding to the best performing models for each case are presented in Table 1. More details regarding these models and the hyperparameters are provided in the Appendix §C.3.

Upon analysis of our experimental results, we observed that direct incorporation of text (T) and image (I) embeddings (Em) as supplementary features to our regression model trained on numeric (N) features resulted in performance degradation. Conversely, when we employed a two-stage approach — first training separate classification models using textual and visual data to generate prediction probabilities (P), then incorporating these probabilities as features in the original regression framework — we achieved significant performance

[12] https://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html (accessed on 8th April, 2025)

improvements. Our methodological workflow is illustrated in the Appendix §C.4 (Figure 2).

Due to constraints in data availability and methodological transparency, comparison with several prior studies was infeasible. Specifically, the models presented in (Qin and Yang, 2019), (Sawhney et al., 2020a), (Sawhney et al., 2020b), and (Sawhney et al., 2021a) could not be replicated, as their implementations rely on audio features which were not included in our dataset. Furthermore, the model proposed in (Medya et al., 2022) is not open source, preventing a comparative analysis.

4.2 Sector Identification & Performance Prediction from Budgets (BASIR)

This study involved two primary experimental components. Firstly, we employed a methodology to identify specific sectors from excerpts of budget transcripts. Secondly, we developed a framework to rank these identified sectors based on their performance, thereby providing a comprehensive analysis of sectoral impacts.

4.2.1 Identifying sectors from excerpts of budget transcripts

The task of identifying sectors from budget excerpts was approached as a multi-class classification problem. We implemented and evaluated several methodologies to address this challenge.

Initially, we employed semantic similarity (STS) based on Nomic embeddings (Nussbaum et al., 2024) to identify sectors from given text excerpts. To enhance performance, we subsequently fine-tuned these embeddings to optimize the vector space representation, such that sectors relevant to a particular excerpt were positioned closer together, while unrelated sectors were distanced. Additionally, we fine-tuned pre-trained language models, specifically BERT (Devlin et al., 2019), and RoBERTa (Liu et al., 2019), for the classification of budget excerpts into appropriate sectors.

The performance metrics for the various models are presented in Table 3. Our analysis reveals that the STS model with fine-tuned embeddings, and τ equals to 0.5 demonstrated superior performance in terms of both Macro (M) and Weighted (W) F1 scores. This suggests that the fine-tuned embedding approach effectively captures the nuanced relationships between budget language and sectoral classifications. Conversely, the BERT model exhibited the highest Micro (m) F1 score.

4.2.2 Ranking Sectors based on their performance

To rank sectors based on their performance, we developed and evaluated four distinct architectural approaches.

Our initial approach involved transforming sector performance data into a binary classification task, determining whether a given sector would experience an upward or downward movement based on the text excerpts related to it. Using this framework, we fine-tuned three encoder-based (Enc) models: BERT (Devlin et al., 2019), RoBERTa (Liu et al., 2019), and DeBERTa (He et al., 2020) for classification purposes. The predicted probabilities from these models were then utilized to generate sector rankings.

Building upon this classification approach, we subsequently fine-tuned the same models for regression analysis. This allowed us to predict the actual performance metrics for each sector with greater precision. The sectors were then ranked according to these predicted performance values, providing a more nuanced assessment of relative sectoral strength.

Following our encoder-based approaches, we implemented feature-based models utilizing Nomic embeddings (Nussbaum et al., 2024) (Emd) extracted from sector-related text excerpts. For binary classification, we trained several machine learning algorithms including logistic regression, random forest, and XGBoost (Chen and Guestrin, 2016). These models were tasked with predicting whether sectors would experience positive or negative performance.

In parallel, we developed regression models using linear regression, random forest, and XGBoost algorithms to predict the actual performance metrics of each sector. The ranking methodology remained consistent with our previous approaches, wherein sectors were ordered based on their predicted performance values. Additionally, we trained an XGBoost model specifically optimized with a learning-to-rank objective to directly produce sector rankings.

In our final experimental approach, we leveraged state-of-the-art large language models (LLMs) to estimate sector performance based on budget text excerpts. Specifically, we employed three advanced LLMs: Gemma-3 27B (Team, 2025), DeepSeek V3 (DeepSeek-AI et al., 2025), and Llama 3.3 70B (Touvron et al., 2023). These models were prompted (§D.3.2) to analyze the sector-relevant text excerpts and estimate the expected performance metrics for each sector. The resulting performance estimates were then utilized to generate sector rankings.

Table 4 presents the comparative performance metrics for these architectural approaches. Notably, the BERT model trained for classification exhibited superior performance in terms of Normalized Discounted Cumulative Gain (NDCG), suggesting that smaller models are more effective when we have a lesser number of instances to train. The performance of the LLMs is comparable to that of the other approaches.

4.3 IPO Rating Prediction from Red Herring Prospectus (BIR)

In this section, we describe the experiments we conducted and discuss the corresponding results.

Following the methodology outlined in (Ghosh et al., 2024), we extracted text from the prospectus (RHP) which were present in PDF format. Optical character recognition (OCR) was performed using Tesseract to extract text from images within the documents. Each page was converted into embeddings utilizing Nomic (Nussbaum et al., 2024). Employing a Retrieval-Augmented Generation (RAG) framework, for each of compiled questions mentioned in Section E.3, we identified the two most pertinent pages based on two criteria: first, through cosine similarity for semantic matching, and second, via BM25 (Lù, 2024) for syntactic similarity. The retrieved pages, along with their corresponding questions, were then passed into the Llama-3.2 3B (AI@Meta, 2024) model to generate answers. Details relating to the prompt we used is mentioned in section E.2. This process yielded a total of 16 answers for each instance, corresponding to the 16 questions posed.

We employed a zero-shot approach by prompting the Gemma-2 9B, Llama 3.1 70B, and Llama-3.2 3B models to classify the aggregate of 16 answers into one of four categories: Apply, May Apply, Neutral, or Avoid. Details of the prompts are provided in section E.2. We then repeated these experiments by substituting the aggregate of answers with a single summary. These summaries were generated using Llama-3.2 3B (AI@Meta, 2024). We observed this change led to improved model performance in most cases. Subsequently, we fine-tuned Llama-3.2 3B and Gemma-2 9B.

Finally, we trained three encoder-based models (RoBERTa (Liu et al., 2019), LongFormer RoBERTa [13], and DeBERTa (He et al., 2020)) with the summaries for classification. The hyperparameters are mentioned in Appendix E.4.

We observed that for MB IPOs, the LongFormer RoBERTa outperformed all other models in terms of micro, macro, and weighted F1 scores. In contrast, for SME IPOs, the Gemma-2 9B model excelled in micro F1 scores, while the Llama 3.1 70B model achieved the highest macro F1 scores. Additionally, the RoBERTa model demonstrated superior performance in terms of the macro FA score. We present the overall flow in Figure 3 and results in Table 5.

5 Conclusion

Our research introduces **InFiNITE**, a comprehensive framework addressing three critical aspects of Indian financial narrative analysis. For corporate earnings calls, our multi-modal approach integrating transcripts, visuals, and market indicators enhances post-announcement stock price prediction accuracy, addressing gaps in traditional single-modality analyses. For Union Budget analysis, we demonstrate that fine-tuned Nomic-based embeddings excel at identifying sectors from budget texts, while BERT-based models effectively rank sectors by predicted performance. This automation enables timely, informed decision-making for investors analyzing budget implications. For IPO evaluation, we present a novel RAG framework that outperforms state-of-the-art LLMs in predicting IPO ratings from prospectuses, supported by specialized datasets for both SME and Main Board listings.

Collectively, these contributions advance computational finance research specifically for the Indian market. Future directions include recommending specific stocks within identified budget-impacted sectors, capturing real-time price movements post-announcements, and developing dynamic question frameworks for red herring prospectus analysis that adapt to industry-specific factors. By bridging NLP with financial expertise, **InFiNITE** establishes a foundation for more sophisticated, data-driven investment decision-making in the Indian context.

[13] `https://huggingface.co/markussagen/xlm-roberta-longformer-base-4096` (accessed on 19th January, 2025)

Limitations

Despite the promising contributions of **InFiNITE**, several limitations must be acknowledged across our three financial narrative analysis tasks.

Data and Sampling Limitations

Our earnings call analysis is restricted to 133 companies representing the Nifty indices, which may not capture the full diversity of the Indian corporate landscape. Our methodology only incorporates instances where both stock price data and comprehensive earnings call materials were available, potentially introducing selection bias.

Similarly, our budget analysis framework emphasized precision over recall in sector identification, with DeepSeek potentially overlooking subtler budget-sector relationships, particularly when policy implications were implicit. This validation approach—focusing exclusively on LLM-detected relationships—potentially reinforces detection bias, creating systematic blind spots in the dataset. Temporal coverage presents significant constraints for budget analysis. Market performance data availability beginning only from 1997 excluded 50 years of budget documents (1947-1996) from complete analysis, limiting insights into long-term policy impacts and historical shifts in sector prioritization. Additionally, inconsistent market data across sectors forced the exclusion of certain sector-period combinations, introducing potential selection bias. To minimize confounding factors, our methodology uses a narrow, immediate event window: the single trading day following the budget announcements and corporate earnings announcements. This aligns with prior research for financial modelling, such as (Sawhney et al., 2021b), (Sawhney et al., 2022), and (Sawhney et al., 2020b).

Methodological Limitations

Due to computational resource constraints, we employed smaller language models rather than state-of-the-art larger models for earnings call analysis, potentially limiting the depth of linguistic understanding. Similarly, for IPO analysis, budget limitations prevented us from using entire prospectuses in PDF format at once. As noted in (Fraga, 2024), larger context sizes can decrease LLM performance and reasoning capabilities, necessitating selective extraction of relevant prospectus sections.

Our IPO analysis utilized a randomized selection of 200 reviews for both MB and SME IPOs, limited by Groq API's free tier rate constraints. We extracted questions using Llama-3 8B (AI@Meta, 2024) and compiled them.

Feature Limitations

Our earnings call analysis does not account for variations in speaking styles, audio data characteristics, or presentation formats, which could contain valuable predictive information beyond textual and visual content.

For budget analysis, our performance metrics isolate budget effects without controlling for confounding macroeconomic factors, sector-specific events, and concurrent corporate announcements that likely influence post-budget market movements. This absence of a comprehensive control framework limits causal interpretations of budget-performance relationships.

References

AI@Meta. 2024. Llama 3 model card.

Ramit Anand and Balwinder Singh. 2019. Effect of composition of board and promoter group retained ownership on underpricing of indian ipo firms: An empirical study. *Indian Journal of Corporate Governance*, 12(1):21–38.

Emanuele Bajo and Carlo Raimondo. 2017. Media sentiment and ipo underpricing. *Journal of Corporate Finance*, 46:139–153.

Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794, New York, NY, USA. ACM.

Xuxia Chen, Jun Wang, and Xi Wu. 2022. Do the outstanding comments of regulatory reviewers for approved ipos serve as a valuation signal for investors? *China Journal of Accounting Studies*, 10(2):147–173.

Saikat Sovan Deb and Vijaya B Marisetty. 2010. Information content of ipo grading. *Journal of banking & Finance*, 34(9):2294–2305.

DeepSeek-AI. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *Preprint*, arXiv:2501.12948.

DeepSeek-AI, Aixin Liu, and Bei Feng et al. 2025. Deepseek-v3 technical report. *Preprint*, arXiv:2412.19437.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of* *the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Sanjay Dhamija and Ravinder Kumar Arora. 2017. Impact of quality certification on ipo underpricing: Evidence from india. *Global Business Review*, 18(2):428–444.

Natanael Fraga. 2024. Challenging llms beyond information retrieval: Reasoning degradation with long context windows. *Preprints*.

Jerome H. Friedman. 2001. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232.

Pierre Geurts, Damien Ernst, and Louis Wehenkel. 2006. Extremely randomized trees. *Machine Learning*, 63(1):3–42.

Sohom Ghosh, Arnab Maji, N Harsha Vardhan, and Sudip Kumar Naskar. 2024. Experimenting with multi-modal information to predict success of indian ipos. *Preprint*, arXiv:2412.16174.

Maarten Grootendorst. 2022. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

H2O.ai. 2025. Distributed random forest (drf). H2O.ai Documentation.

Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2020. Deberta: Decoding-enhanced bert with disentangled attention. *arXiv preprint arXiv:2006.03654*.

T. K. Huynh and V. Shenai. 2019. Option trading volumes and their impact on stock prices at earnings' announcements: A study of s&p100 stocks in the post crisis era 2010-2017. *International Journal of Academic Research in Accounting, Finance and Management Sciences*, 9(3):83–103.

Joshy Jacob and Sobhesh Kumar Agarwalla. 2015. Mandatory ipo grading: does it help pricing efficiency? *Vikalpa*, 40(2):132–144.

Mrunal Joshi and Rucha Mehta. 2018. Impact of union budget on stock market. *Contemporary Issues in Marketing and Finance*, 1:29–45.

Zahid Younas Khan, Zhendong Niu, Sulis Sandiwarno, and Rukundo Prince. 2021. Deep learning techniques for rating prediction: a survey of the state-of-the-art. *Artificial Intelligence Review*, 54:95–135.

Zahid Hassan Kharuri, T Manjunatha, and V Rajesh Kumar. 2021. Stock price reactions to budget announcement in indian capital market. *International Journal of Science and Management Studies*, 4(6):59–69.

Rohit Kumar, Sourabh Bikas Paul, and Nikita Singh. 2024. Words that move markets-quantifying the impact of rbi's monetary policy communications on indian financial market. *arXiv preprint arXiv:2411.04808*.

Xiaojiang Lei, Xueming Qian, and Guoshuai Zhao. 2016. Rating prediction based on social sentiment from textual reviews. *IEEE transactions on multimedia*, 18(9):1910–1921.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach.

Tim Loughran and Bill McDonald. 2011. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of finance*, 66(1):35–65.

Xing Han Lù. 2024. Bm25s: Orders of magnitude faster lexical search via eager sparse scoring. *Preprint*, arXiv:2407.03618.

T Manjunatha and Zahid Hassan Kharuri. 2023. Effects of budget announcement on stock prices in the indian context. *Asian Journal of Management*, 14(1):57–64.

A Mansurali, P Mary Jayanthi, R Swamynathan, and Tanupriya Choudhury. 2022. Social listening on budget—a study of sentimental analysis and prediction of sentiments using text analytics & predictive algorithms. In *Machine Intelligence and Data Science Applications: Proceedings of MIDAS 2021*, pages 879–892. Springer.

Geo Martin. 2024. Analyzing the impact of the union budget on sectoral indices in the national stock exchange (nse).

Sourav Medya, Mohammad Rasoolinejad, Yang Yang, and Brian Uzzi. 2022. An exploratory study of stock price movements from earnings calls. In *Companion Proceedings of the Web Conference 2022*, WWW '22, page 20–31, New York, NY, USA. Association for Computing Machinery.

Meta AI. 2025. Llama 4: The beginning of a new era of natively multimodal intelligence. Meta AI Blog. Https://ai.meta.com/blog/llama-4-multimodal-intelligence/.

Zach Nussbaum, John X. Morris, Brandon Duderstadt, and Andriy Mulyar. 2024. Nomic embed: Training a reproducible long context text embedder. *Preprint*, arXiv:2402.01613.

Vivek Panwar and Ganesh Kumar Nidugala. 2019. Impact of budget and gdp announcements on indian stock market. *Finance India*, 33(4):929–946.

Sanjay Poudyal. 2008. *Grading Initial Public Offerings (IPOs) in India's Capital Markets: A Globally Unique Concept*. Indian Institute of Management.

Yu Qin and Yi Yang. 2019. What you say and how you say it matters: Predicting stock volatility using verbal and vocal cues. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 390–401, Florence, Italy. Association for Computational Linguistics.

Lizhen Qu, Georgiana Ifrim, and Gerhard Weikum. 2010. The bag-of-opinions method for review rating prediction from sparse text patterns. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 913–921, Beijing, China. Coling 2010 Organizing Committee.

Seshadev Sahoo and Prabina Rajib. 2010. After market pricing performance of initial public offerings (ipos): Indian ipo market 2002–2006. *Vikalpa*, 35(4):27–44.

Vishal Sarin and Neeru Sidana. 2017. A study of perceptions of investors towards ipo grading in india. *International Journal of Economic Research*, 14(20):757–770.

Ramit Sawhney, Arshiya Aggarwal, Piyush Khanna, Puneet Mathur, Taru Jain, and Rajiv Ratn Shah. 2020a. Risk forecasting from earnings calls acoustics and network correlations. In *INTERSPEECH*, pages 2307–2311.

Ramit Sawhney, Arshiya Aggarwal, and Rajiv Ratn Shah. 2021a. An empirical investigation of bias in the multimodal analysis of financial earnings calls. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3751–3757, Online. Association for Computational Linguistics.

Ramit Sawhney, Mihir Goyal, Prakhar Goel, Puneet Mathur, and Rajiv Ratn Shah. 2021b. Multimodal multi-speaker merger & acquisition financial modeling: A new task, dataset, and neural baselines. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 6751–6762, Online. Association for Computational Linguistics.

Ramit Sawhney, Piyush Khanna, Arshiya Aggarwal, Taru Jain, Puneet Mathur, and Rajiv Ratn Shah. 2020b. VolTAGE: Volatility forecasting via text audio fusion with graph convolution networks for earnings calls. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8001–8013, Online. Association for Computational Linguistics.

Ramit Sawhney, Megh Thakkar, Ritesh Soun, Atula Neerkaje, Vasu Sharma, Dipanwita Guhathakurta, and Sudheer Chava. 2022. Tweet based reach aware temporal attention network for NFT valuation. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 6321–6332, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Model	Modalities	MAE	RMSE	MAPE
DL-1	N	150.769	269.193	**0.288**
DL-2	N+ T (Em)	228.321	348.152	0.454
DL-3	N+ T (P)	125.204	216.639	0.349
DL-4	N+ T (Em) + I (Em)	271.350	457.369	0.965
DL-5	N+ T (P) + I (P)	**104.787**	**188.537**	0.334
Llama-4	N + T (Raw) + I (Raw)	108.417	246.196	5.918

Table 1: Results. Details of the models are mentioned in §C.3. Deep Learning (DL), Numeric (N), T (Text), I (Image), Embedding (Em), Predicted Probabilities (P)

Anshul Saxena, Vandana Vijay Bhagat, and Amrita Tamang. 2021. Stock market trend analysis on indian financial news headlines with natural language processing. In *2021 Asian Conference on Innovation in Technology (ASIANCON)*, pages 1–5. IEEE.

Gemma Team. 2025. Gemma 3.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. Llama: Open and efficient foundation language models. *Preprint*, arXiv:2302.13971.

Domonkos F Vamossy. 2025. Social media emotions and ipo returns. *Journal of Money, Credit and Banking*, 57(1):31–67.

A Reproducibility

The datasets, codes, and documentation can be accessed from `https://github.com/annonymous-upload/resources-2025`

B Tables

All the tables referred to in the paper are presented here.

C MiMIC: Appendix

C.1 MiMIC: Annotation decision rationale

To ensure a comprehensive and representative sample of the Indian equity market, the dataset incorporates firms from the Nifty 50 (large-cap), Nifty MidCap 50 (mid-cap), and Nifty SmallCap 50 (small-cap) indices. This stratified selection mitigates potential biases associated with an exclusive focus on large, widely analyzed corporations, thereby enhancing the generalizability of findings. Screener.in aggregates publicly available financial data and earnings call documents. Therefore, any bias inherent in Screener.in's coverage would largely reflect the publicly disclosed information landscape for listed Indian companies. The

Metric	Budget Transcripts	Sector Identification	Sector Ranking
Total Entries	97	1,671	429
Temporal Span	1947–2025	1947–2025	1997–2025

Table 2: Dataset Statistics

	F1 (M)	F1 (m)	F1 (w)
STS (base)	0.159	0.176	0.345
STS (fine-tune)	**0.291**	0.478	**0.605**
BERT	0.179	**0.489**	0.425
RoBERTa	0.075	0.274	0.192

Table 3: Results of Multi-Label Sector Classification

Model	Type	NDCG
BERT	**Enc Clasifier**	**0.997**
RoBERTa	Enc Clasifier	0.994
DeBERTa	Enc Clasifier	0.996
BERT	Enc Regressor	0.995
RoBERTa	Enc Regressor	0.995
DeBERTa	Enc Regressor	0.995
Logistic	Emd + Classifier	0.996
Random Forest	Emd + Classifier	0.996
XG-Boost	Emd + Classifier	0.994
Linear	Emd + Regressor	0.995
Random Forest	Emd + Regressor	0.996
XG-Boost	Emd + Regressor	0.994
XG-Boost	Learning to Rank	0.994
Gemma-3 27B	Zero Shot	0.994
DeepSeek V3	Zero Shot	0.993
Llama 3.3 70B	Zero Shot	0.994

Table 4: Sector Ranking Results

dataset is built upon this publicly available information, mirroring the data accessible to a general investor or analyst. We acknowledge that any data source may have subtle inherent biases, but the company selection process was designed to counteract a narrow focus. Owing to computational limitations, the present analysis is constrained to 133 listed firms; nonetheless, the underlying methodology is adaptable and may be readily extended to a broader cohort by including additional constituents from the target indices. To further reduce the influence of confounding variables, the study employs a narrowly defined event window, limited to the single trading day immediately following each corporate earnings announcement.

C.2 MiMIC: Details of Numeric Data

C.2.1 Macroeconomic Variables:

Gross Domestic Product (GDP) Growth, Inflation Rate

C.2.2 Market Data:

NIfty 50 Opening Price, Nifty 50 Closing Price, Nifty 50 Volume

C.2.3 Technical Indicators:

Simple Moving Averages (SMA20, SMA50), Relative Strength Index (RSI14)

C.2.4 Fundamental Indicators:

A comprehensive set of fundamental variables was collected for each company. Due to the annual frequency of this data, we utilized the previous year's values for training and prediction. **Financial statement items** (Sales, Expenses, Operating Profit, Other Income, Interest Expense, Depreciation, Profit Before Tax, Tax Rate, Net Profit, EPS, Dividend Payout, Equity Capital, Reserves, Borrowings, Other Liabilities, Total Liabilities, Fixed Assets, CWIP, Investments, Other Assets, Total Assets),
Cash flow items (Cash from Operating Activities, Cash from Investing Activities, Cash from Financing Activities, Net Cash Flow),
Additional metrics (Revenue, Financing Profit, Financing Margin, Deposits, Borrowing)

		MB			SME		
Model	**Input**	**F1 (m)**	**F1 (M)**	**F1 (w)**	**F1 (m)**	**F1 (M)**	**F1 (w)**
Gemma-2 9B (Zero Shot)	All Answers	0.009	0.007	0.005	0.411	0.189	0.368
Llama-3.1 70B (Zero Shot)	All Answers	0.039	0.021	0.054	0.374	0.176	0.355
Llama-3.2 3B (Zero Shot)	All Answers	0.484	0.184	0.348	0.076	0.038	0.114
Gemma-2 9B (Zero Shot)	Summary	0.023	0.108	0.012	**0.516**	0.256	0.416
Llama-3.1 70B (Zero Shot)	Summary	0.115	0.044	0.191	0.457	0.281	**0.423**
Llama-3.2 3B (Zero Shot)	Summary	0.162	0.077	0.255	0.429	0.163	0.361
Llama 3.2 3b (SFT)	Summary	0.836	0.228	0.883	0.361	0.299	0.347
Gemma 2 9B (SFT)	Summary	0.716	0.233	0.814	0.402	0.298	0.349
RoBERTa	Summary	0.769	0.219	0.846	0.406	**0.335**	0.377
LongFormer RoBERTa	Summary	**0.968**	**0.246**	**0.952**	0.224	0.126	0.090
DeBERTa	Summary	0.912	0.239	0.925	0.457	0.319	0.383

Table 5: Model Performances. m = micro, M = Macro, w = weighted, SFT = Supervised Fine-tuning. Best performing models are highlighted in bold.

C.3 MiMIC: Hyper-parameters

The hyper-parameters of the models discussed in this paper, are presented here.

C.3.1 Text Embedding based classifier

Model Type: XGBoost
Number of trees: 30

C.3.2 Image Embedding based classifier

Model Type: Distributed Random Forest
Number of trees: 40
minimum depth: 13, maximum depth: 20
minimum leaves: 94, maximum leaves: 115

C.3.3 Regression Model

Model Type: Feed-forward based neural network (DL-5), Number of layers: 3, Number of hidden units: 20, Dropout: 10

Hyper-parameters of other models (i.e., DL-1 to DL-4) and other information in detail are provided in the code base.

C.4 MiMIC: Workflow

Our methodological workflow is illustrated in Figure 2.

C.5 MiMIC: Prompt

You are an expert financial analyst. Using the earnings call transcript, images from the presentation slides, technical indicators, macroeconomic variables, market data, fundamental indicators, and the opening price on the earnings release day, estimate the opening stock price of the company on the day next to the day of the earnings call. Only provide the answer as a real number. No need for any justification.
Input Text: <*text along with tables in markdown format*>
Input Numeric: <*numeric data along with column names in json format*>
Input Images: <*list of input images*>

D BASIR: Appendix

D.1 BASIR: Annotation decision rationale

Due to budgetary constraints, our annotation pipeline for the BASIR dataset used DeepSeek for

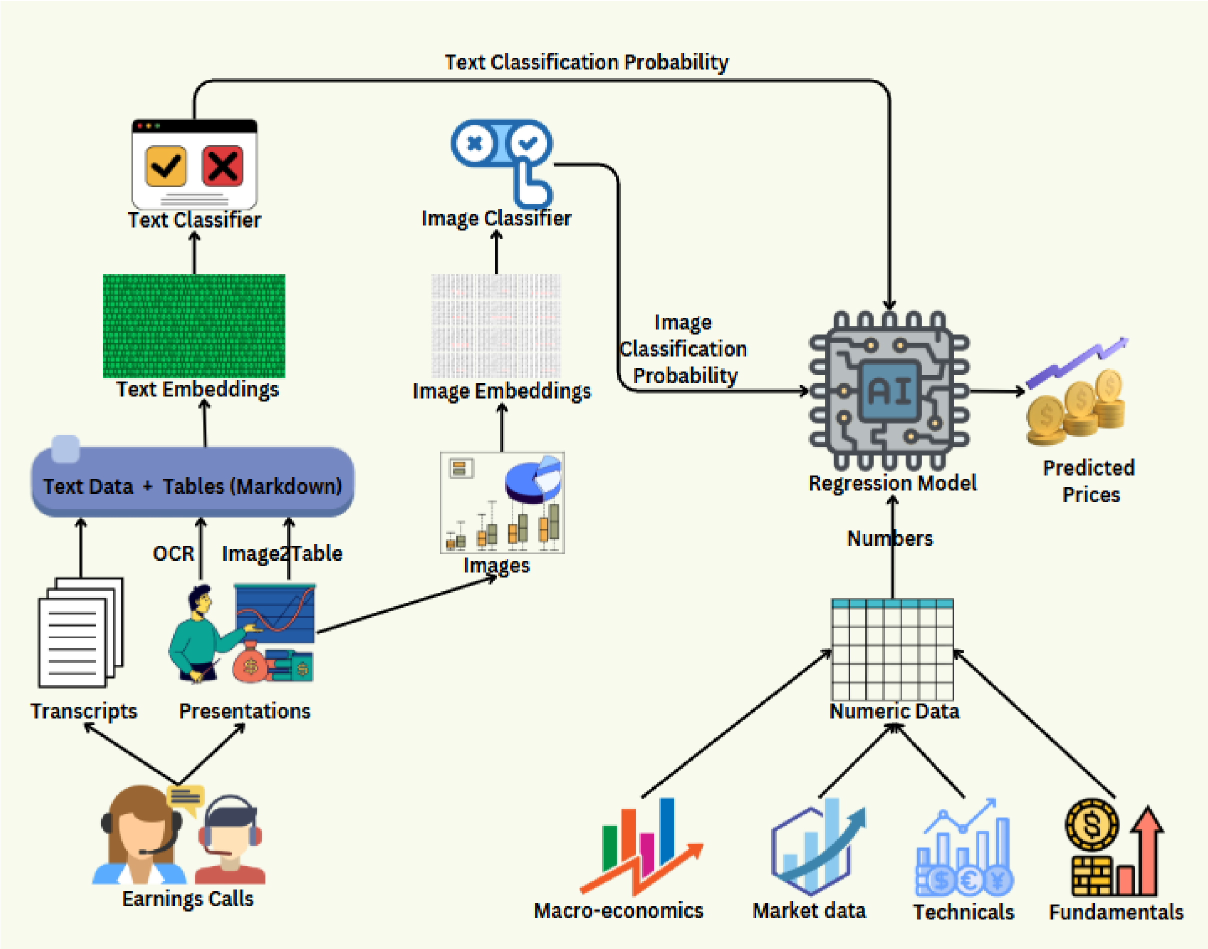

Figure 2: Workflow

pre-annotation, followed by a 100% manual validation by a single financial industry expert with over five years of experience. This "expert-in-the-loop" approach ensures high consistency across the dataset. The expert's primary task was to correct errors and discard any LLM hallucinations, ensuring the final data's reliability. Because a single expert established the ground truth, inter-annotator agreement is not applicable, while data consistency is maximized.

D.2 BASIR: Industries

List of industries are as follows: ['Aerospace & Defence' , 'Agro Chemicals' , 'Air Transport Service' , 'Alcoholic Beverages' , 'Auto Ancillaries' , 'Automobile' , 'Banks' , 'Bearings' , 'Cables' , 'Capital Goods - Electrical Equipment' , 'Capital Goods-Non Electrical Equipment' , 'Castings , Forgings & Fastners' , 'Cement' , 'Cement - Products' , 'Ceramic Products' , 'Chemicals' , 'Computer Education' , 'Construction' , 'Consumer Durables' , 'Credit Rating Agencies' , 'Crude Oil & Natural Gas' , 'Diamond , Gems and Jewellery' , 'Diversified' , 'Dry cells' , 'E-Commerce/App based Aggregator' , 'Edible Oil' , 'Education' , 'Electronics' , 'Engineering' , 'Entertainment' , 'Ferro Alloys' , 'Fertilizers' , 'Finance' , 'Financial Services' , 'FMCG' , 'Gas Distribution' , 'Glass & Glass Products' , 'Healthcare' , 'Hotels & Restaurants' , 'Infrastructure Developers & Operators' , 'Infrastructure Investment Trusts' , 'Insurance' , 'IT - Hardware' , 'IT - Software' , 'Leather' , 'Logistics' , 'Marine Port & Services' , 'Media - Print/Television/Radio' , 'Mining & Mineral products' , 'Miscellaneous' , 'Non Ferrous Metals' , 'Oil Drill/Allied' , 'Packaging' , 'Paints/Varnish' , 'Paper' , 'Petrochemicals' , 'Pharmaceuticals' , 'Plantation & Plantation Products' , 'Plastic products' , 'Plywood Boards/Laminates' , 'Power Generation & Distribution' , 'Power Infrastructure' , 'Printing & Stationery' , 'Quick Service Restaurant' , 'Railways' , 'Readymade Garments/ Apparells' , 'Real Estate Investment Trusts' , 'Realty' , 'Refineries' , 'Refractories' , 'Retail' , 'Ship Building'

, ‘Shipping’ , ‘Steel’ , ‘Stock/ Commodity Brokers’ , ‘Sugar’ , ‘Telecomm Equipment & Infra Services’ , ‘Telecomm-Service’ , ‘Textiles’ , ‘Tobacco Products’ , ‘Trading’ , ‘Tyres’]

D.3 BASIR: Prompts

D.3.1 Text Extraction and Sector Identification

You are provided with the budget of India below. From this budget only pick up text segments relevant to the given list of industries. List of industries: <list of industries> Your output should be a json file having 2 keys: ‘text_segment’ and ‘industry’. The value corresponding to ‘text_segment’ would be the extract text segment extracted from the budget. The value of ‘industry’ should be the corresponding list of industries from the given list that the text segment is related to. Return only the segments having any relation with the given list of industries. One text segment can be related to multiple industries.

Text context from Budget: <Budget Transcript of a given year>

D.3.2 Sectorwise Performance Prediction

You are a financial expert with extensive experience of analysing Indian Budgets. Given a sector and an excerpts related to the sector from a budget speech, estimate the performance of the sector. You output should be just a real number between -1 to 1. Don’t reply anything else. Sector: <name of sector>, Excerpt: <text excerpts related to the given sector>

E BIR: Appendix

E.1 BIR: Annotation decision rationale

For BIR, the trustworthiness of our labels is empirically validated. We established a strong correlation between the expert recommendations used as labels and the subsequent financial performance of the IPOs. We reveal that an “Apply” recommendation from an expert reviewer corresponded with a positive listing-day return in 82.17% of MainBoard IPOs and 83.49% of SME IPOs. This demonstrates a direct, quantifiable link between the labels and real-world market outcomes. Our methodology is consistent with established financial research like (Chen et al., 2022), and (Vamossy, 2025) which shows that expert opinions and investor sentiment act as significant signals for IPO valuation and performance. By sourcing reviews from expert analysts at reputable firms, we ensure the labels are not arbitrary but are reliable proxies for an IPO’s prospective success.

E.2 BIR: Prompts

Question Extraction Prompt:
The prompt used for extracting questions is:
You are an expert financial analyst who have extensive experience of participating in Initial Public Offerings (IPOs) of Indian companies. You are given a review about an Indian company going for IPO. Extract a list of key questions which have been answered in the given review and which would help in determining whether to apply for the IPO. Return just a list of questions which can be answered from the review. Do not return anything other than the list of questions. Review: {review content}
Response:

Answer Generation Prompt:
This prompt was used for each of the 16 questions to generate the corresponding answer.
You are an expert financial analyst who have extensive experience of participating in Initial Public Offerings (IPOs) of Indian companies. Relevant contents from Red Herring Prospectus (RHP) of an Indian company going for IPO is given to you. Your task is to analyse and answer the given question in less than 300 words as free text. Use just the content provided to you to answer the question and not anything else. If the contents are not relevant, just return the word ‘None’.
CONTENT-1: {semantically relevant content }
CONTENT-2: {syntactically relevant content}
Question: {question}
Response:

Summary Generation Prompt:
The prompt used for generating summary from answers is as follows:
You are an expert financial analyst who have extensive experience of participating in Initial Public Offerings (IPOs) of Indian companies. You are provided with various facts about a company going for IPO in the form of answers. Your task is to analyse these answers and generate a summary comprising of key points that investors needs to know to decide if they should subscribe for the IPO or not. If you are not confident answer nan. Just return the summary in 300 words and nothing else. Facts about the company’s IPOs are as follows: {answers of 16 questions}.

Response:

Rating Prediction Prompt:
The prompt used for zero shot classification is:
"You are an expert financial analyst who has extensive experience of participating in Initial Public Offerings (IPOs) of Indian companies. You are given various facts of a company. Your task is to analyse these facts and decide whether an investor should 'Avoid', 'May apply', 'Apply', or, be 'Neutral' for the IPO. Your answer should be in a JSON structure with two keys, 'prediction' and 'justification'. The value corresponding to 'prediction' key should be 0,1,2, or, 3 only where 0 represents 'Avoid', 1 represents 'Neutral', 2 represents 'May apply', and 3 represents 'Apply'. The value corresponding to 'justification' key should be the explanation behind the prediction. Facts: {answers of 16 questions concatenated side by side}.
Response:"

E.3 BIR: Questions

We needed to identify key sections in the prospectus that would best inform IPO ratings. To accomplish this, we randomly selected 200 reviews each from MB and SME IPOs. We then processed these selected reviews through the Llama-3 8B model, extracting questions using the prompt outlined in Section §E.2. This process yielded a consolidated list of 16 unique questions. The list of questions is presented here.

- What is the price band and issue price of the IPO?
- What is the issue size and how many shares are being issued as part of the IPO?
- What is the implied market capitalization of the company after the IPO?
- How will the company utilize the funds raised through the IPO, and what is the purpose of the IPO?
- What is the company's revenue growth rate over recent financial years, and how has its financial performance been historically (including revenue, EBITDA, and net profit trends)?
- What are the key financial ratios, such as net profit margin, return on equity (RoE), return on capital employed (RoCE), and total debt?
- What is the shareholding pattern before and after the IPO, and who are the promoters?
- Are there any regulatory issues or conflicts of interest affecting the company?
- What are the company's plans for expansion and future growth, and how does it position itself in terms of competition within its industry?
- Who are the company's major customers, what is the revenue breakdown by sector, and is there a dependency on large institutional customers?
- What are the potential risks associated with increasing raw material costs, and what other risks does the company face?
- How does the company's valuation compare to its peers, and is the issue priced aggressively compared to industry standards?
- What is the competitive landscape of the industry in which the company operates?
- Has the company declared any dividends in the past, and what is its dividend policy?
- Who are the lead managers and registrar for the IPO, and what is their track record in terms of past IPO listings?
- Are there any concerns regarding transparency or missing details in the offer document?

E.4 BIR: Hyper-parameters

Encoder based models
learning_rate=2e-5, per_device_train_batch_size=1, per_device_eval_batch_size=1, num_train_epochs=5, gradient_accumulation_steps=4, weight_decay=0.01

Decoder based models
max_seq_length = 204, load_in_4bit = True, lora_alpha = 16, lora_dropout = 0, bias = "none", use_gradient_checkpointing = "unsloth", random_state = 3407, use_rslora = False, dataset_num_proc = 2, packing = False, per_device_train_batch_size = 2, gradient_accumulation_steps = 4, warmup_steps = 5, num_train_epochs=5, learning_rate = 2e-4, optim = "adamw_8bit", weight_decay = 0.01, lr_scheduler_type = "linear"

E.5 BIR: Workflow

F Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used perplexity.ai in order to improve readability and language of the work. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

G Potential Risks

The datasetes has been released under the CC-BY-NC-SA-4.0 licence for non-commercial research purposes only. We are not liable for any monetary loss that may arise from the use of these datasets and model artifacts.

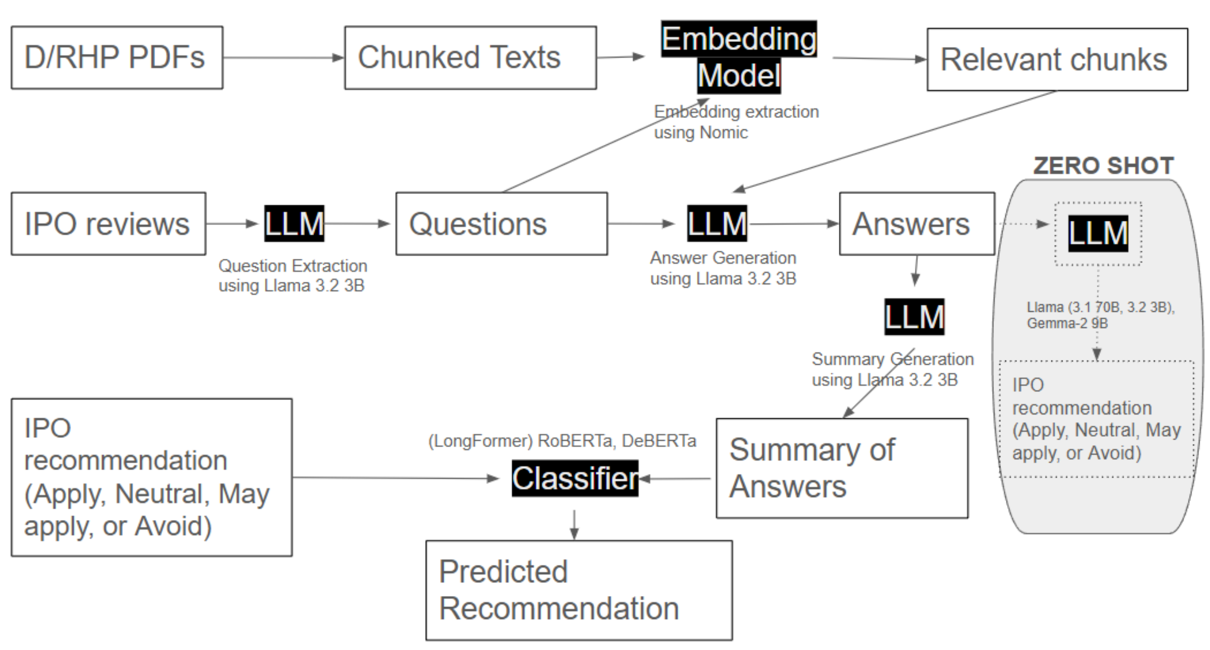

Figure 3: Detailed Flowchart narrating the predicting ratings of Indian IPOs