

Intelligent Regulatory Insight System for Indian Public Companies: An AI Driven Financial Forensics and Compliance Platform

Swapna Lokhande¹, Rishikesh Koli², Tejashwini Patil³, Isha Sonkusare⁴ and Yashraj Kulkarni⁵

¹⁻⁵Department of Computer Science and Business Systems, St. Vincent Pallotti College of Engineering & Technology, Nagpur, India

Emails: swapna.lokhande@gmail.com, 1983rishikesh@gmail.com, tejashwinipatil875@gmail.com, ishasonkusare.006@gmail.com, yashrajulkarni0405@gmail.com

Abstract— The INTELLIGENT REGULATORY INSIGHT SYSTEM on Indian public companies is a Multi Agent Financial Forensics Application that assists in enhancing Fraud Detection, Risk Assessment and Regulatory Compliance of Indian listed companies. Such system is a combination of structured financial statements, Stock market data, News information, social media sentiment analysis, and enforcement information to form a synchronized 10 Agent system framework. The system is able to give natural language form explanations with support by context and evidence material using the transformer language models and Retrieval Augmented (RAG) frameworks.

Experiments have been done on a representative sample of large cap Indian companies, and the results show that the system consistently provides data integration and normalization, financial distress early warnings, and narratives that are easy to understand and use. As a whole, the system design ensures auditability and scalability of the solution in regulatory settings.

Index Terms— Financial Forensics, Indian Capital Markets, SEBI, Fraud Detection, FinBERT, Regulatory technology, Multi Agent Systems, Risk Assessment.

I. INTRODUCTION

However, for public equity markets to function effectively, financial reporting must be consistent and truthful. Earnings manipulation, hidden liabilities, delayed reporting and other similar practices cause a lack of balance of information available to all stakeholders.[1] This, in turn, makes it necessary for regulatory bodies to intervene in order to protect investors once damage has been done.[2, 3]

In India, increasing scale and sophistication of publicly traded entities make it increasingly complicated for the regulators to oversee the process. Although traditional auditing techniques can be helpful, they will prove to be inefficient in assessing large amounts of both structured and unstructured data continuously produced by modern financial market environments.[3,4,5] Financial reports, trading activity, media analysis, expert assessments and regulatory decisions combine into a complicated system of information that needs more advanced tools for efficient processing.[5]

It is evident that there is an urgent need for development of automated systems capable of structuring and analyzing data on financial indicators, textual sources, and comparisons with peers.

[5, 6] They will provide valuable assistance to experts in detecting inconsistencies and summarizing risk factors in a clear manner.

The Intelligent Regulatory Insight System for Indian publicly traded companies can be considered as such. The system will increase the supervisory capability by adding Multi Source Data intake, forensic analytics, composite risk scoring and transformer based question answering, without compromising understandability and responsibility. [6, 7]

II. LITERATURE REVIEW

A. Financial Forensics and Red-Flag Approaches

These are the Traditional methods of Forensic Accounting, which involves the use of Ratio analysis, checks trends and the use of heuristic red flags.

These methods are non-secretive but relatively small and lacking in variety and tend to ignore agenda and market cues.[2]

B. Financial Language Models

FinBERT domain tuned pre trained models are useful to enhance sentiment and classification of financial text.[1, 4] The semantic retrieval of data is supported by embeddings of such models, and this is essential to the evidence linked explanations in our system.

C. Regulatory Technology and Supervisory Technology

To Supervise the accuracy of information we use, the RegTech and SupTech systems, which encompasses surveillance tools, transactions monitoring as well as automated disclosure checks.[5, 6] Nevertheless, systems that are used to integrate in-depth analysis of financial statements, text interpretation, and peer benchmarking are still being developed and will be finished shortly.

III. METHODOLOGY

A. System Architecture and Agents

The Intelligent Regulatory Insight System is designed using a modular microservice architecture built around 10 Co-ordinated Agents (Table 1). The system relies on FastAPI to handle external request, while Celery and Redis support asynchronous task execution and workflow management. Structured financial records are stored in PostgreSQL, ensuring reliable data organization and retrieval. In addition, a vector database is used to maintain Document Embeddings for semantic search and questions and answering tasks.[3] Each agent is assigned a specific responsibility within the overall workflow, allowing the system to remain scalable, updatable, and adaptable to future extensions.

- Agent 1: Data Ingestion
- Agent 2: Forensic Analysis
- Agent 3: Risk Scoring
- Agent 4: Compliance Validation
- Agent 5: Report Generation
- Agent 6: Orchestration
- Agent 7: Question & Answer Interface
- Agent 8: Market Intelligence
- Agent 9: Peer Benchmarking
- Agent 10: Regulatory Monitoring

B. Data Sources and Normalisation

This system aggregates data from a wide variety of sources to ensure complete information. For financial information, sources like Yahoo Finance, along with the data collected from NSE and BSE are considered. The use of Financial Modeling Prep is considered for more detailed historical data. PDF data released by companies, SEBI circulars, news data and Google Trends data are also aggregated as part of this process.

Documents that may be required to be aggregated through OCR are cleaned to prepare them before further processing. This ensures that all structured data can be brought into a standard internal schema. This system is also capable of tracking any sources of data and inconsistencies in them as part of data cleaning processes.

TABLE I: AGENTS USED IN THE INTELLIGENT REGULATORY INSIGHT SYSTEM FOR INDIAN PUBLIC COMPANIES

Agent	Role	Main Inputs and Outputs
1	Data Ingestion	Inputs: Yahoo Finance, NSE, BSE, FMP, filings, news. Output: Normalized tables in PostgreSQL and text in ChromaDB.
2	Forensic Analysis	Input: normalized statements. Output: twenty nine forensic metrics with quality flags.
3	Risk Scoring	Input: Metrics. Output: six dimensional risk scores and contributing factors
4	Compliance Validation	Inputs: Metrics, regulatory report. Output: potential non compliance Indicators
5	Report Generation	Inputs: Metrics, Scores, Flags. Output: PDF reports, dashboards, CSV Exports
6	Orchestration	Inputs: Request for analysis. Output: Schedules, status and logs.
7	Q&A Interface	Inputs: User questions. Output: answers while referring to documents.
8	Market Intelligence	Inputs: News, Google Trends, Prices. Output: Knowing the pulse of the moment.
9	Peer Benchmarking	Inputs: Metrics, sector data. Output: Peer distributions, z-scores and Outlier lists.
10	Regulatory Monitoring	Inputs: Enforcement orders, circulars. Output: Company links and timelines.

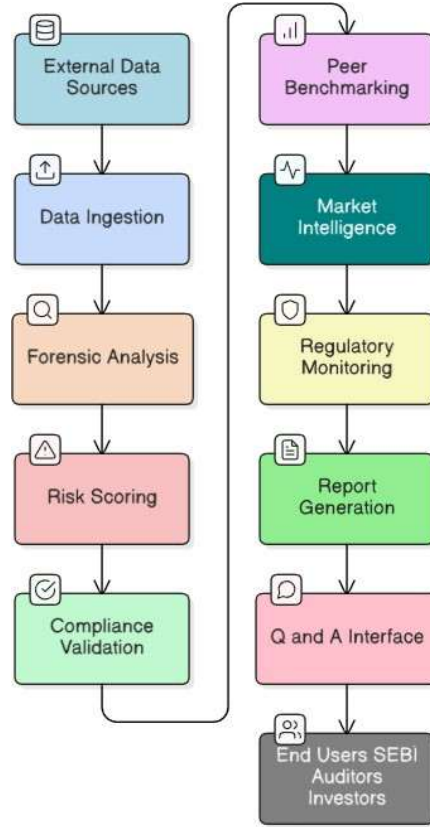


Figure 1. Processing pipeline of the Intelligent Regulatory Insight System for Indian Public Companies, from external data sources through analysis agents to end users such as SEBI analysts, auditors, and investors.

C. Forensic Metrics

Agent 2 computes 29 metrics in three groups:

Vertical analysis (11): common-size shares (revenue share, COGS share, operating expense share, operating income share, net income share, current assets share, fixed assets share, intangible assets share, current liabilities share, long-term debt share, equity share).

- Horizontal analysis (10): QoQ/YoY growth rates for revenue, operating income, net income, assets, liabilities, operating cash flow, receivables, inventory, cash.
- Financial ratios (8): current ratio, quick ratio, debt-to-equity, interest coverage, ROA, ROE, gross margin, operating margin.

Each metric carries a quality flag: primary-source, imputed, or conflict.

D. Risk Dimensions and Compliance Checks

Six risk dimensions: earnings quality, leverage & solvency, liquidity pressure, revenue stability, governance & disclosure, and market behaviour anomalies. Continuous metrics are standardized within peer groups as z-scores and combined with binary indicators:

$$R = \sum_{i=1}^n w_i z_i + \sum_{j=1}^m v_j I_j$$

Weights w_i and v_j are domain-informed and tunable.

Agent 4 performs arithmetic reconciliations and disclosure searches like related party disclosures and changes in auditors, with alerts raised where mandated text is absent or when there are close matches.[5] In addition, differences in the equation Assets = Liabilities + Equity or discrepancies in the cash flow statement with the notes on the account will automatically raise a High level alert.

IV. EXPERIMENTS AND RESULTS

A. Experimental Setup

Initial evaluation on a curated set of 50 large cap Indian companies across sectors like Information Technology, Finance, Manufacturing from 2016-2024.

Objectives: Intake consistency, early warning system, runtime performance, and Report quality.

Environment: Commodity server (Intel i7, 32GB RAM), PostgreSQL, Redis, Celery. Embeddings via FinBERT style encoder.[1]

B. Data Quality and Consistency

Across 3,200 company quarterly financial reports:

- 92.4% of numeric fields matched across sources with accuracy close to 2%.
- 5.7% differences due to consolidation or timing.
- 1.9% flagged for manual review (restatements or OCR's).

C. Composite Risk Scoring Evaluation

Ground-truth: 14 distressed events (2018–2024). Metrics:

TABLE II: RISK SCORING EVALUATION

Metric	Score
Precision @ 10	0.70
Recall @ 10	0.64
Mean Risk (Distressed)	0.72
Mean Risk (Healthy)	0.31
Separation Margin	0.41

A. Case Study 1: Early Warning

The manufacturing company showed a steady increase in both, i.e., liquidity and leverage risk approximately four quarters prior to its restructuring event. The liquidity score increased from 0.29 to 0.74, while leverage risk rose from 0.23 to 0.69. A noticeable increase in governance risk was also observed following the resignation of the external auditor (Table 3). Further analysis indicated rising receivables, declining operational cash flows and persistently negative sentiment, which together contributed to the elevated risk assessment generated by our Intelligent Regulatory Insight System.

B. Case Study 2: Earnings Quality

A consumer company-oriented revenue growth of 9% year by year, while receivables increased by 48% over the same time period. At the same time, operating cash flow declined by 12% year by year and news sentiment

remained negative. Based on this combination of indicators, the system classified the Company as having high earnings quality risk (Table 2). Subsequent Regulatory Scrutiny was consistent with this assessment.

C. Performance Benchmarking

Comparison Vs Single process baseline:

TABLE III: PERFORMANCE COMPARISON

Task	Baseline	INTELLIGENT REGULATORY INSIGHT SYSTEM	Speedup
Ingestion + Metrics (1 company, 4 quarters)	16.2 s	4.8 s	3.3×
Risk Score Computation	1.1 s	0.28 s	4×
LLM Q&A Response	3.6 s	1.9 s	1.9×

D. Expert Evaluation

The Three domain experts rated narrative outputs:

- Clarity: 4.6/5
- Evidence linkage: 4.2/5
- Usefulness: 4.4/5
- Interpretability: 4.7/5

E. Limitations

The Areas of work for future are The Limited labelled enforcement events, restricted XBRL coverage and initial weight settings require supervised calibration.

V. CONCLUSION

Intelligent Regulatory Insight System was designed as a Multi Agent system that combines structured financial analysis, market intelligence, regulatory monitoring, and LLM supported explanations to strengthen supervisory capabilities for Indian public companies (Figure 1). By combining information from multiple sources, the platform is able to identify potential early warning signals and generate clear, understandable reports that experts can use for initial assessment and decision making.

Further, The future development of the project will emphasize expanding XBRL compatibility, enhancing scoring weight calibration using historical enforcement data, developing related party identification using graph models, and assessing the system in regulated settings. Through these improvements, the Intelligent Regulatory Insight System can be an effective aid for regulators, exchanges, auditors, and investors in ensuring transparency and maintaining market integrity.

ACKNOWLEDGMENT

We the authors of this Research paper would like to thank the domain experts and reviewers who provided feedbacks and test cases for the Research Paper and for our knowledge too. This work used publicly available financial datasets and regulatory documents for evaluation.

REFERENCES

- [1] D. Araci, "FinBERT: Financial Sentiment Analysis with Pre trained Language Models," arXiv:1908.10063, 2019
- [2] S. Jaiswal, "Deep Learning-based Anomaly Detection in Financial Statements," IEEE Access, 2022.
- [3] B. Zhang, "Large Language Models for Regulatory Compliance Automation," Proc. ACM FAccT, 2023.
- [4] A. H. Huang et al., "FinBERT: A Large Language Model for Extracting Information from Financial Text," 2023.
- [5] R. Goyal, "AI-driven Supervisory Technology for Capital Markets: A Survey," Journal of Financial Regulation and Compliance, 2023.
- [6] Securities and Exchange Board of India, "Memorandum of Understanding between SEBI and National Forensic Sciences University," 2025.
- [7] Securities and Exchange Board of India, "Empanelment of Agencies for Providing Digital Forensics Services," 2025.