
Automated SEBI Filings Analysis Using BERT for Financial Compliance

[GitHub Report]

By Sibi Marappan

Internship Period: July – Dec 2022

Company: HCL

Duration: 6 months

LinkedIn: [Sibi Marappan on LinkedIn](#)

Table of Contents:

1. [Introduction](#)
 2. [Agenda and Purpose](#)
 3. [Key Features and Outputs](#)
 4. [Why Not Use GPT to Summarize Data?](#)
 5. [Questions and Answers](#)
 6. [Who Can Use This Project?](#)
 7. [Future Enhancements](#)
 8. [Conclusion](#)
-

1. Introduction

This project automates the analysis of regulatory documents issued by the **Securities and Exchange Board of India (SEBI)** using **Natural Language Processing (NLP)**, **Transformer-based models** (specifically **BERT**), and **Machine Learning** techniques. The system processes SEBI filings to extract critical financial and regulatory information, categorize documents by topic, and visualize relationships between financial entities and regulations.

The project utilizes advanced data cleaning and text preprocessing, along with **named entity recognition (NER)**, **topic modeling**, and **graph-based visualization** to provide a comprehensive understanding of SEBI filings. The fine-tuned **BERT model** used in this project is specialized to understand SEBI-specific language, making it more effective than generic NLP models.

This project was part of my **HCL Internship**, with a focus on addressing real-world regulatory challenges faced by financial institutions.

2. Agenda and Purpose

The main goals of this project are:

- **Automating the analysis** of SEBI's financial filings to save time and reduce the manual burden of compliance officers.
 - **Providing insights** through the extraction of key terms, topic classification, and visualizations of financial relationships within SEBI documents.
 - Helping compliance officers, auditors, and financial analysts make faster, more accurate decisions by streamlining the document review process.
-

3. Key Features and Outputs

Entity Extraction

The system extracts key financial entities such as company names, insider trading actions, penalties, and regulatory terms. For example:

- **Company X, insider trading, penalty, compliance officer.**

Topic Categorization

Each document is categorized by its dominant topic (e.g., **insider trading, mergers, compliance**). The system assigns relevance percentages to show how strongly a document relates to a particular topic.

Summarization

Automatic summaries provide users with a quick overview of key actions or penalties discussed in SEBI filings. Example:

- "This document outlines SEBI's decision on insider trading by Company X."

Graphical Visualization

Interactive graphs show relationships between companies, regulations, and penalties. For instance:

- A visual graph showing **Company X** connected to insider trading regulations and further linked to penalties, helping users see patterns and connections.
-

4. Questions and Answers

Q1: What is this project about?

This project automates the analysis of SEBI's regulatory filings using NLP and machine learning. It extracts critical terms, categorizes topics, and provides a visual overview of relationships between financial entities and regulations.

Q2: How will someone use this project?

Users upload SEBI filings to the system. The tool processes them by extracting key terms, categorizing the document by topics (e.g., insider trading, compliance), generating summaries, and visualizing connections between entities.

Q3: Why not use Google or GPT for this?

While Google can find documents and GPT can summarize, neither provides the level of financial and regulatory specificity that this project does. This project fine-tunes BERT for SEBI-specific terms and allows for **batch processing** and **relationship visualization**, capabilities not offered by Google or GPT.

Q4: Don't I still need Google to download new SEBI filings?

Yes, you would need Google or the SEBI website to download filings. However, once the documents are downloaded, this project automates their analysis and provides insights that would take hours if done manually.

Q5. Why Not Use GPT to Summarize Data?

While **GPT** can summarize text, it is a **general-purpose tool**. This project uses a **fine-tuned BERT model** trained specifically on SEBI filings, providing more accurate and specialized results:

1. **Tailored for SEBI:** This BERT model understands legal and financial terms in SEBI regulations, making it more precise than GPT.
 2. **Batch Processing and Visualization:** This project can analyze multiple documents simultaneously and visualize the relationships between entities, which GPT cannot do.
 3. **Specialized Entity Extraction:** The model extracts relevant terms like company names, insider trading, and penalties with a level of accuracy GPT lacks.
-

5. Who Can Use This Project?

1. **Compliance Officers:**
 - Ensures companies comply with SEBI regulations by quickly identifying insider trading violations and penalties.
 2. **Financial Analysts:**
 - Analyzes the impact of new regulations on companies and industries by reviewing trends across SEBI filings.
 3. **Legal Experts:**
 - Extracts legal terms and decisions for advising clients on financial regulations.
 4. **Auditors:**
 - Identifies compliance issues and risks in multiple SEBI filings, particularly those related to insider trading or penalties.
-

6. Future Enhancements

1. **Automated SEBI Filings Download:**
 - Integrating a web scraper or API to automatically download the latest SEBI filings will eliminate the need for manual downloads.

2. Dashboard for Non-Experts:

- Developing a user-friendly dashboard that simplifies insights for non-experts and presents visual overviews without technical knowledge.

3. Real-Time Monitoring:

- Real-time monitoring could continuously analyze new SEBI filings and deliver instant alerts for critical terms (e.g., insider trading or penalties).

4. Customizable Reports:

- Users could generate customized reports on specific topics (e.g., insider trading or compliance violations) based on their preferences.

7. Conclusion

This project offers a powerful, automated solution for analyzing SEBI filings. By combining **fine-tuned NLP models**, **machine learning techniques**, and **graphical visualizations**, it transforms complex regulatory documents into actionable insights for compliance officers, auditors, and financial analysts. The project's ability to process multiple documents at once, extract key financial entities, and provide topic categorization and relationship visualization makes it a valuable tool for ensuring regulatory compliance.

By automating the tedious manual review process, this project enables professionals to focus on higher-level decision-making, helping them stay up to date with the latest SEBI regulations in a fast and efficient way.
