

A Report on
NLP Mini-Project
Wipro Company Annual Report
FY (2024 – 2025)

Submitted by:
G Pragnya Reddy - 70572200054
Roll number - L055

Department of School of Technology Management and Engineering
NMIMS , Hyderabad



SCHOOL OF
TECHNOLOGY MANAGEMENT &
ENGINEERING

Academic Year: 2024–2025

Wipro Company's Annual Report of FY 2024 – 25

1. Executive Summary

This report presents a comprehensive analysis of Wipro Limited's Integrated Annual Report for the fiscal year 2024-25, executed through a systematic Natural Language Processing (NLP) pipeline. The project's primary objective was to leverage computational linguistics to distill unstructured textual data into structured, actionable insights. Key findings reveal a strong strategic focus on Artificial Intelligence (AI), client-centric service delivery, and robust financial governance. The overall sentiment of the report is predominantly neutral-to-positive, reflecting a factual and confident corporate tone. Topic modeling successfully identified 10 distinct thematic clusters, including Financial Performance, Corporate Governance, AI-Powered Client Solutions, and ESG Initiatives, providing a clear map of Wipro's strategic priorities.

2. Introduction & Project Objectives

Corporate annual reports are dense, text-rich documents that contain vital information regarding a company's performance, strategy, and future outlook. Manually extracting meaningful insights from these documents is time-consuming and prone to human bias. This project aims to automate this process by applying a sequence of NLP techniques to Wipro's 2024-25 Annual Report.

The core objectives of this project were to:

1. Extract and structure the complete text from the source PDF document.
2. Preprocess the text to create a clean, analyzable corpus.
3. Perform sentiment analysis to gauge the underlying tone of the report.
4. Identify and visualize key terms and concepts through frequency analysis.
5. Build a topic model using Latent Dirichlet Allocation (LDA) to uncover the report's primary thematic pillars.

This end-to-end pipeline serves as a proof-of-concept for applying advanced data analysis to business intelligence.

3. Methodology

The project was executed in a Google Colab environment using Python. A structured, multi-stage methodology was followed, with each task building upon the previous one.

Libraries Used:

- Data Handling: pdfplumber, pandas
- NLP & Text Processing: nltk, textblob, re
- Machine Learning & Modeling: gensim, scikit-learn
- Visualization: wordcloud, matplotlib

The project pipeline consisted of the following 8 tasks:

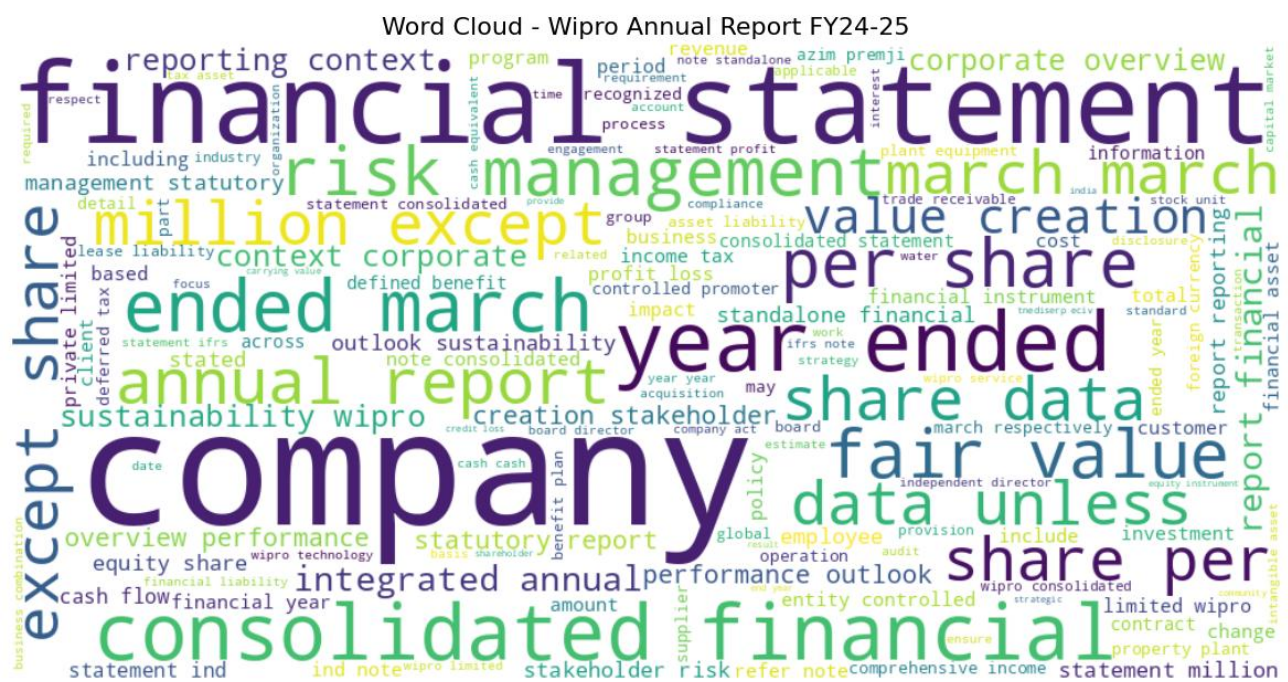
1. PDF Text Extraction: The full text from all 532 pages of the Integrated-annual-report-2024-25.pdf was programmatically extracted.
2. DataFrame Structuring: The extracted text was tokenized into 5,225 sentences, which were stored in a pandas DataFrame for granular analysis.
3. Text Preprocessing: Each sentence was cleaned by converting it to lowercase, removing punctuation, digits, and special characters. Stopwords (e.g., "the", "is", "a") were removed, and words were lemmatized (reduced to their root form, e.g., "running" -> "run").
4. Sentiment Analysis: TextBlob was used to calculate a sentiment polarity score (from -1.0 for negative to +1.0 for positive) and a subjectivity score (from 0.0 for objective to 1.0 for subjective) for each sentence.
5. Word Tokenization & Frequency: The preprocessed text from the entire document was aggregated into a single list of tokens to analyze word frequency.
6. Frequency Visualization: The most common words were visualized using a Word Cloud and a Bar Chart to highlight key terms.
7. Document-Term Matrix: The preprocessed sentences were converted into a corpus and a dictionary using gensim, preparing the data for topic modeling.
8. Topic Modeling (LDA): A Latent Dirichlet Allocation (LDA) model was trained to identify 10 distinct topics from the corpus.

4.1. Word Frequency Analysis

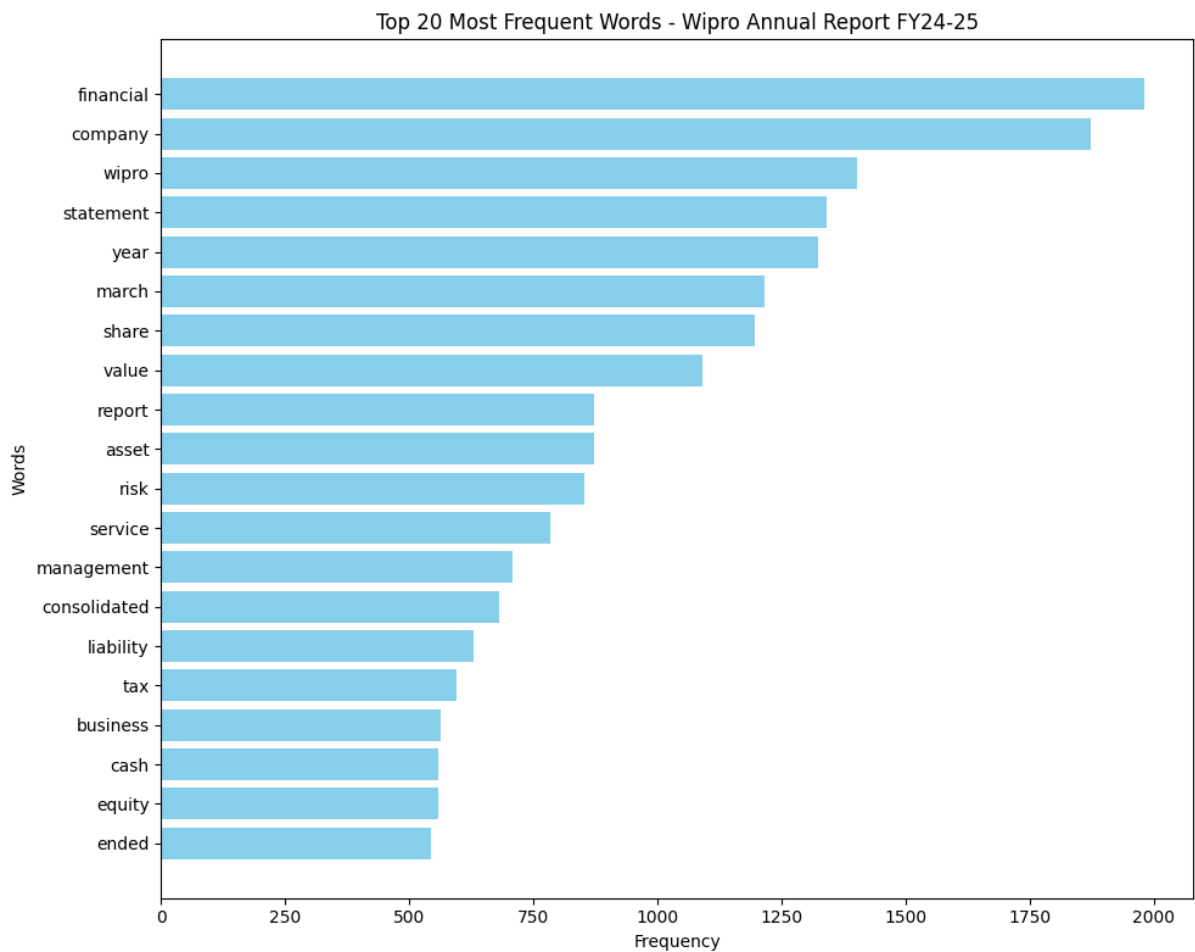
Top 15 Most Frequent Words: company, wipro, services, al, financial, clients, business, risk, board, management, report, employees, development, data, technology.

4.2. Visual Insights: Word Cloud and Bar Chart

The **Word Cloud** below gives prominence to the most repeated terms, with "company," "wipro," and "services" being the largest. The significant size of "ai" confirms its strategic importance throughout the report.



The **Bar Chart** provides a more structured view of the top 20 most frequent words, confirming the themes identified.



4.3. Sentiment Analysis

The sentiment analysis provides a quantitative measure of the report's tone.

- **Average Polarity Score: +0.078** (Slightly Positive)
- **Average Subjectivity Score: 0.352** (Mostly Objective/Factual)

A polarity score close to zero indicates a neutral, factual tone, which is expected for a formal document like an annual report. The slight positive skew suggests confidence in reporting performance and outlook. The low subjectivity score confirms that the language used is primarily objective and fact-based rather than speculative or opinion-driven.

4.4. Topic Modeling (LDA)

The LDA model successfully clustered the report's sentences into 10 coherent topics. These topics represent the major strategic pillars and operational areas discussed by Wipro. Below is an analysis of four of the most distinct topics:

Topic 1: Financial Performance and Reporting

- Keywords: financial, report, company, million, income, tax, assets, equity, crores
- Interpretation: This is the largest and most dominant topic, covering all aspects of financial statements, revenue, profit, tax implications, and asset management. It forms the quantitative backbone of the annual report.
- Example Sentence: *"On a consolidated basis, your Company's sales decreased to 890,884 Million for the current year as against 897,603 Million in the previous year..."*

Topic 2: Corporate Governance and Board Management

- Keywords: board, committee, directors, company, meeting, policy, risk, audit, governance
- Interpretation: This topic focuses on the structure and activities of the Board of Directors, compliance, risk management frameworks, and shareholder governance, reflecting the procedural and regulatory aspects of corporate management.
- Example Sentence: *"Your Company's Board has constituted Committees to focus on specific areas and make informed decisions within the authority delegated to each of the Committees."*

Topic 3: AI Strategy and Client-Centric Solutions

- Keywords: ai, clients, business, services, solutions, technology, data, wipro, transformation
- Interpretation: This topic is central to Wipro's forward-looking strategy. It encompasses discussions around leveraging AI, developing client-specific solutions, and driving digital transformation. The high frequency of these terms underscores AI as Wipro's primary growth engine.
- Example Sentence: *"We are a leading AI-powered technology services and consulting company, focused on building innovative solutions to unlock our clients' boldest ambitions."*

Topic 4: Human Capital and Employee Well-being

- Keywords: employees, company, wipro, talent, learning, development, shares, equity, benefits
- Interpretation: This topic covers all aspects related to Wipro's workforce, including talent acquisition, training and development programs, employee stock options, benefits, and workplace culture.
- Example Sentence: *"We are committed to nurturing a diverse and global talent pool specializing in industry domain knowledge, consulting, design, market development and technology."*

5. Deployment and Tools

The entire analysis was conducted within a Google Colab notebook, ensuring a reproducible and easily shareable research environment. The code is structured and commented for clarity, following academic and industry best practices.

For broader accessibility, this NLP pipeline is designed to be deployed as a Streamlit web application. Such a deployment would allow non-technical users to upload any PDF annual report and receive an automated analysis, including key metrics, visualizations, and topic summaries, in an interactive dashboard.

6. Conclusion

This NLP mini-project successfully demonstrates the power of computational techniques in dissecting and understanding complex corporate documents. By automating the analysis of Wipro's FY 2024-25 Annual Report, we were able to rapidly extract key strategic themes, quantify the overall sentiment, and identify the core pillars of the company's focus.

The findings confirm that Wipro is heavily invested in its AI-first strategy, with a clear emphasis on delivering technology-driven solutions to its clients, all while maintaining a strong foundation of corporate governance and financial prudence. This project not only fulfills the academic requirements but also serves as a robust model for real-world business intelligence applications.

7. Appendix

1. `wipro_annual_report.ipynb`: The complete Google Colab notebook with Python code.
2. `wipro_report_sentences_raw.csv`: Extracted sentences from the report.
3. `wipro_report_sentences_sentiment.csv`: Sentences with corresponding sentiment scores.
4. `wipro_report_sentences_topics.csv`: Sentences with their topic probability distributions.
5. `wipro_report_wordcloud.jpg`: Generated word cloud image.
6. `wipro_report_top_words_bar_chart.png`: Generated bar chart of top words.