Company Info & Financial Analyzer - Technical Report

Approach:

This application allows users to analyze a company by name. It uses a language model to generate a set of related URLs and scrapes content from those URLs using BeautifulSoup and Playwright.

Model Selection:

We selected Google's Gemini model (`gemini-1.5-flash-latest`) via the LangChain framework due to its performance in reasoning, summarization, and instruction following. The model is used to:

- Generate relevant URLs

- Summarize scraped content

- Perform financial/fundamental analysis

Data Preprocessing:

We perform URL validation to exclude social media and scrape pages using BeautifulSoup for static content and Playwright for JavaScript-heavy pages. We clean HTML by removing tags like `<script>`, `<style>`, `<nav>`, etc., and extract only meaningful `<body>` text.

Performance Evaluation:

We evaluate model performance based on:

- Quality of URL generation

- Completeness and accuracy of scraped data

- Relevance and coherence of generated summaries

Cited Model:

Gemini 1.5 Flash (via `langchain-google-genai`)

This application does not use a static dataset but dynamically fetches content per user input.