Pre-project Task: Variable Recommendation System in Credit Report

Task introductions — Tue 9 Sep 2pm

Wed 10 Sep 10:30 am — Task release

Task submission deadline — Wed 24 Sep 10:29 am
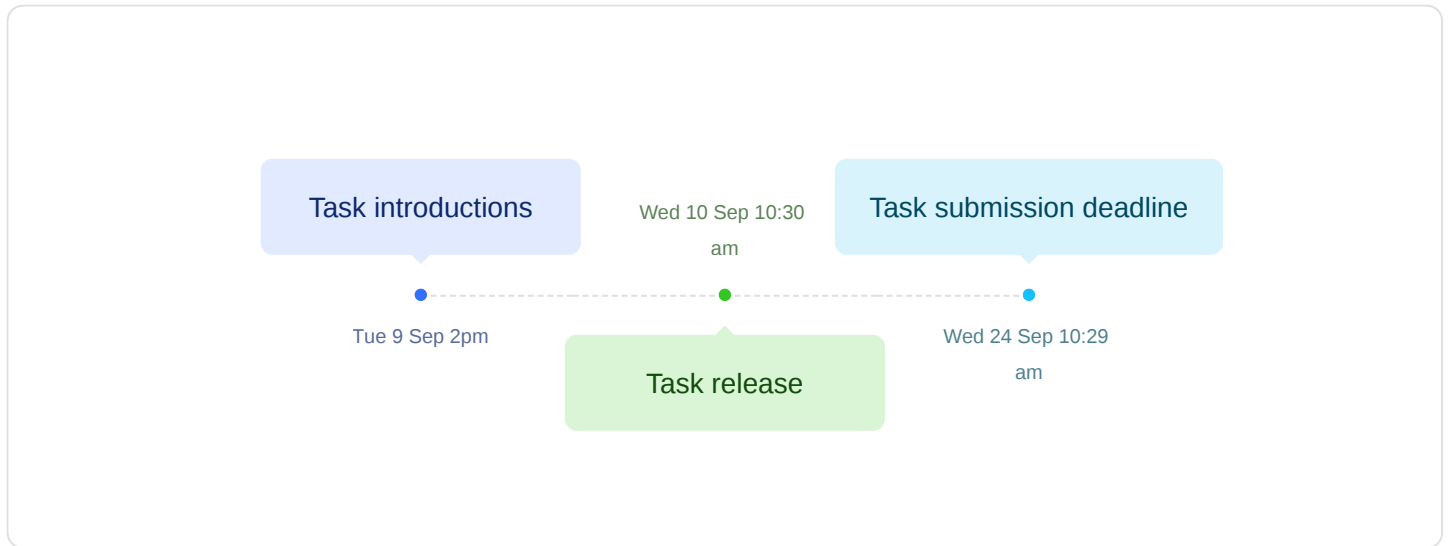
Background

Analysts writing or updating subsections of corporate credit rating reports often need a quick, defensible list of **variables/events/factors** that typically matter for a given company and subsection (e.g., *Liquidity*, *Leverage & Coverage*, *Earnings/Volatility*, *Regulatory/Legal*, *ESG*, etc.). We aim to build a recommender that suggests such variables — grounded in historical Fitch rating reports and provides **traceable evidence** (sentences/paragraphs) for each suggestion.

Task Description:

The ultimate goal is to figure out a way to automatically identify important news and other relevant factors for credit rating analysis, leveraging a recommendation system approach. As a pre-project task, the scope is simplified. That is, given a company name and a (sub)section name, figure out a way to propose relevant variables, events, or factors that could be considered when writing that subsection. This can involve methods inspired by collaborative filtering or other recommendation techniques. Moving forward, the news containing those variables, events or factors could be identified as important news.

Using Fitch Rating reports and referencing the report subsections from our database, the candidate should generate a set of suggested variables for each subsection that could guide analysis or reporting.

Objective

The primary objective is to develop a prototype system that:

1. Identifies and ranks key variables/events/factors relevant to a company's credit rating subsection.

2. Demonstrates the feasibility of applying recommendation system methodologies in financial report analysis.

Data Provided:

1. Moody's report in HTML

2. A sample of subsection name and description pairs (encouraged to extend this list with additional subsections)

Please see the email for the link to dropbox

Focus on credit rating report subsections such as Business Risk, Financial Risk, and other key areas relevant to Fitch Rating analysis.

Deliverables

1. **Source Code**:
   - A clean, well-documented codebase in a zipped folder.
   - Scripts for each stage of data processing, analysis, and final output.
   - `README.md` detailing the project's purpose, methodology, data sources, and instructions for running the code.
   - A `requirements.txt` or `pyproject.toml` specifying all necessary libraries and dependencies.

2. **Final Results**:
   - A sample output demonstrating the recommended variables/events/factors for selected companies and report subsections.

3. **Process Documentation**:
   - **Methodology**: A step-by-step explanation of the approach taken, such as data preprocessing, modeling techniques, and ranking strategies.
   - **Challenges and Solutions**: Discussion of any challenges encountered during the task and how they were addressed.

- **Optional Analysis**: Insights on trends, patterns, or interesting observations derived from the data.

Timeline & Submission

- **Timeline**: The candidate has **2 weeks** from the date of receipt to complete this task.

- **Submission**: Please upload your codes and results to **Google drive** and send **the google link via email** as submission. **(Please make sure everyone with a link can access it.)**