Guage - Financial sentiment analysis App

I utilized a large number of analyst reports, each focusing on a different stock. Leveraging RAG, NLP, and an LLM, the project objective was to predict sentiments accurately. The resulting app allows users to easily input reports and gain insights into whether stocks are worth buying, should be sold, etc. The deployment was successfully achieved using Flask.

**Implementation Steps:**

Extract Text from Analyst Report PDF'S:

* Developed a Python script to systematically extract text from each analyst report.
* Identified and extracted crucial sections like ‘Highlights’, ‘Investment Rationale’, and ‘Industry Outlook’, as deemed important by annotators.

Keyword Extraction:

* Employed NLP techniques such as lowercasing, tokenization, stopword removal, punctuation removal, non-alphabetic token removal, and stemming for text preprocessing.
* Generated a new CSV containing keywords for each PDF.

Compiling A Test Dataset:

* Utilized RAG by including a retrieval model to fetch relevant external data, and combined this information with the analyst reports to annotate and form a test CSV containing the out-performance of all stocks based on sentiment, conducted by a human annotator.

Integrating and Finetuning LLM:

* Passed the keywords to a pretrained LLM (DistilRoBERTa, via HuggingFace) to predict sentiments.
* Addressed large text issue by finetuning the model, using ‘chunk-size’ approach.
* Retrieved the predicted sentiments and evaluated accuracy by comparing these retrieved sentiments with the test CSV.

Deployment on Flask:

* Integrated all modules properly for seamless functionality.
* Designed the webpage structure and styling using HTML and CSS.
* Users can input text of any length or drop a PDF directly onto the Flask UI to receive sentiment output.