FSIL Application: Targeted LLM-Inference

Yahya Hassan

Introduction

For my application to FSIL, my project is an interactive visualization that can provide clear insight into stock relationships and relative performance of S&P 500 listed companies through the use of clustering and emotion detection. A variety of tools have previously been used to try and represent stock market performance. My intent in this project is to add another dimension to stock analysis through emotional detection in 10-K filings using an efficient form of LLM analysis.

Problem Definition

In the finance domain, particularly stock analysis, there is an abundance of data and underutilization of Natural Language Processing (NLP) models' underlying value of emotion processing. The sheer volume of data can be overwhelming and computationally expensive to process, especially when employing Large Language Models (LLMs) which are resource intensive. The financial industry relies on precise and timely data-driven decisions that require efficient data analysis methods. High computational costs and timely delivery is a necessity. Even the best financial analysis reports can quickly become stale if they are late or come at a high computational cost, especially in markets with narrow windows for high-profit actions. The preliminary step of data preprocessing has become increasingly crucial to allow efficient handling and analysis of large datasets. Refining data before it enters NLP models reduces processing times, lowers cost, and makes sophisticated language technologies more accessible and effective. While there are many data preprocessing techniques, such as feature engineering, PCA, and LDA, and masking strategies to make LLMs more efficient, there are currently very few tools that combine multiple different metrics of financial analysis into the preprocessing of data and even fewer that provide flexibility and transparency into the process. NLP models, especially those based on deep learning, often operate as "black boxes," providing little to no insight into how decisions are derived. Even though they have achieved state-of-the-art performance in tasks such as translation and summarization that can only be slightly improved by human reinforcement, tasks such as market sentiment analysis are more likely to see more significant improvement when human intuition is integrated. Additionally, inherent randomness in predictions, from random initial weights, further complicates them.

Proposed Method

The status quo of finding which stocks have underlying value is that there really isn't one. Institutions assign millions of dollars to their brightest minds to try and capitalise on undervalued assets consistently, but the task remains difficult. The proposed solution involves a two-stage process. Split into a dynamic clustering visualization and NLP model that draws inference from the refined subset passed in from the clustering.

A. Clustering Algorithm and Visualization. Firstly, the visualisation and clustering algorithm. Using a mixture of Gaussian mixture models and K-means clustering, the backend algorithm detects financial patterns using the latest available data provided by the Yahoo Finance API. The team was able to clearly visualise and highlight potentially undervalued companies to shorten the analysis process. In order to demonstrate the findings of the algorithms, an involved data communication process was used between front-end and back-end systems. Multiple local APIs were created to communicate the data being transferred and modules such as flask and npm were used extensively to provide data for the main visualisation in d3.

The methodology of the analysis followed three key steps. Firstly, the data collection and

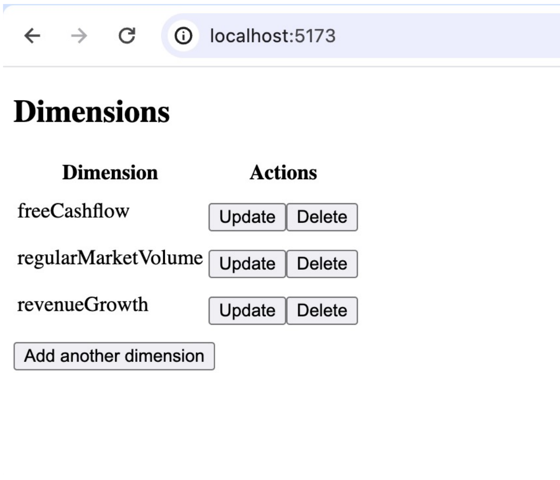

Fig. 1. Example dimensions set by user

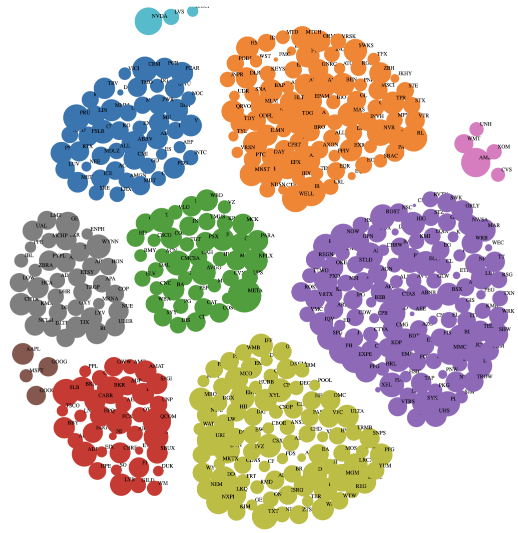

Fig. 2. The finished visualization clusters on default settings (zoomed out).

clustering back-end algorithms extract the financial data from Yahoo Finance's API using basic requests and cluster it according to the stock features selected by the user. Membership of clusters is evaluated by an expectation-maximization(EM) algorithm that creates a cluster membership fitting into a Gaussian distribution with user-defined features.

Secondly, the process of connecting the front end and back ends of the system. To ensure standardized, structured, and reliable data exchange between the front end and back end, a Flask API communication protocol was used. Using RESTful principles, the front end and back end were able to communicate dynamically. Therefore, the application can leverage Python's powerful processing capabilities and the results seamlessly return to the user. One of the key advantages of using RESTful principles in Flask is the standardization of data exchange. By standardizing exchange to JSON, a lightweight data-interchangeable format, the complexities of both systems are abstracted to only a handful of endpoints. This approach simplifies the development process by introducing modularity to the application's design. Flask integration allowed rapidly prototype, develop, and deploy new features as the project is developed iteratively. This use of API communication additionally ensured routing to specific function files would stay reliable even as folders were shifted around in the submission process. Since these routings are only relative, they did not get affected by the main folder being placed into the final submission folder.

Lastly, the construction of the graphical illustration showing the clusters. Using the financial data processed from the back-end, the visualization first renders a drop-down menu dynamically based on the cluster values, then organizes ticker nodes by cluster and creates links between ticker nodes within each cluster, using an SVG initialization to specify width and height. Visualization updates dynamically based on selected cluster from the drop-down menu. These nodes are created using specific sizing metrics, to represent the average value of each stock-feature combination's deviation away from the cluster mean. What this allows the user to do is to identify any companies that are clustered in a group that could be an outlier, and could therefore be incorrectly priced in the market.

Overall, this subsection of the solution allows the user to input metrics of interest that guide this algorithm to create clusters of related stocks to make NLP inference more efficient and relevant.

B. NLP inference. As input, the NLP inference aspect of this project takes in a list of companies within a cluster and the financial metrics of interest.

The stock tickers of the companies are then used to access the 10-K Filings through the EDGAR database. Using the a expanded definition of each item typically on a 10-K filing against user-defined financial metrics, attention scores are drawn through a pretrained FinBERT model. Using FinBert is more efficient since it is trained on financial jargon and gives more relevant attention scores. Attention scores are compared to a threshold that helps narrow down which items from the 10-K are more insightful and likely to drive investment decisions. This saves compute resources and minimizes the randmoness in NLP inference.

Items of relevance from 10-K filings from stock tickers of companies within the cluster are then passed into two pipelines. Firstly, a emotion classification pipeline. Distilbert-base-uncased-emotion classifies emotion into anger, disgust, fear, joy, sadness, and surprise. Finbert is also an alternative that classifies emotion into positive, neutral, and negative. Distilbert, however, may be better to use simply for the purpose of increasing the variety of information available to the user. Emotional sentences within the items of relevance seperated by each ticker are found and passed into the summarization pipeline. The summarization pipeline uses a t5-base model that tokenizes and summarizes sentences while maintaining context.

```
{'anger': ['Antitrust, Unfair Competition, and Overcharge Class Actions  Antitrust and unfair
competition class action lawsuits were filed against us in British Columbia, Ontario, and
Quebec, Canada.',
  'Product-Related Litigation  U.S.',
  'We assumed responsibility for these claims in our agreement to acquire Nokia’s Devices and
Services business and have been substituted for the Nokia defendants.',
  'The plaintiffs allege that their handsets either operated outside the FCC Guidelines or
were manufactured before the FCC Guidelines went into effect.',
  'The lawsuits also allege an industry-wide conspiracy to manipulate the science and testing
around emission guidelines.',
  'The plaintiffs have filed supplemental expert evidence, portions of which the defendants
have moved to strike.',
```

Fig. 3. Sample ’anger’ sentences from Microsoft 10-K filings

Experiments and Evaluation

The experiments aim to address several key inquiries: Firstly, evaluate the accuracy of the model and identify areas for improvement, particularly in determining the significance of features influencing market pricing. Secondly, the usefulness and effectiveness towards the user in informing decision.

Below is a list of fundamental questions, associated experiments to address them, and descriptions of those experiments:

1. **How can emotion detection be more effective at financial analysis than financial sentiment analysis?** Emotion detection adds a more nuanced interpretation of market sentiments compared to traditional

```
TICKER: AAPL
  ITEM: Item 6
Your max_length is set to 100, but your input_length is only 76. Since this is
a summarization task, where outputs shorter than the input are typically
wanted, you might consider decreasing max_length manually, e.g. summarizer('...
', max_length=38)
JOY: the information set forth below for the five years ended September 26,
2020, is not necessarily indicative of results of future operations .

  ITEM: Item 7
ANGER: the Company's senior management has reviewed these critical accounting
policies and related disclosures with the Audit and Finance Committee of the
Company’s Board of Directors .

FEAR: the Company is subject to income taxes in the U.S. and numerous foreign
jurisdictions . resolution of uncertainties in a manner inconsistent with
management’s expectations could have a material impact on the Company’s
financial condition and operating results . extent of the continuing impact of
the COVID-19 pandemic is uncertain and will depend on many factors outside the
Company's control .

JOY: the company’s net income for the fiscal year ended September 24, 2018 was
$1,057 million . the company's net sales for the year ended september 30, 2018
were $1,217 million - a 4% increase over the same period last year . a net loss
for the quarter ended sept. 30, 2018 was $47 million, compared to the same
quarter last year, which was $57.4 billion .

SADNESS: weakness in foreign currencies relative to the U.S. dollar had a
significantly unfavorable year-over-year impact on Rest of Asia Pacific net
sales during 2022 . iPhone iPhone net sales decreased 2% or $4.9 billion during
2023 compared to 2022 due to lower net sales of non-Pro iPhone models .

  ITEM: Item 8
ANGER: in the opinion of management, there was not at least a reasonable
possibility the Company may have incurred a material loss . VirnetX filed a
lawsuit against the Company alleging that certain of the Company’s products
infringe on patents owned by virnetX . the company disagrees with the FCA’s
decision, and has appealed .

FEAR: Apple Inc. is subject to taxation and files income tax returns in the U.
s. federal jurisdiction and many state and foreign jurisdictions . because of
its inherent limitations, internal control over financial reporting may not
prevent or detect misstatements . the Company has not recognized revenue, and
has elected not to disclose amounts, related to these undelivered services .
```

Fig. 4. Sample summary of sentences according to emotion from APPL

financial sentiment analysis by capturing a broader range of human emotions. The enhanced granularity provides a new dimension of financial analysts with valuable insights into the underlying investor behaviors and market dynamics. Emotion detection plays to the strength of NLP models. NLP models are more effective and well-trained to suit the complexities of natural language, enabling them to identify subtle variations in tone and context that accompany different emotions. This allows for a more nuanced interpretation of text, which can be crucial for understanding the full spectrum of sentiments expressed in financial documents and communications. By analyzing these varied emotions, analysts can gain a deeper understanding of investor sentiment, potentially predicting market movements with greater accuracy. For instance, fear might indicate potential sell-offs, while joy could suggest bullish market trends. That accompanied with financial analysis can increase investor confidence.

   By identifying emotional trends, emotion detection aids in better risk management, highlighting potential overvaluations or undervaluations that may not be evident through standard analysis. Additionally, it enhances communication strategies by aligning marketing efforts with the prevailing emotional climate, ultimately leading to more effective engagement with investors and customers. Overall, emotion detection integrates psychological and sociological perspectives into market analysis, offering a comprehensive view that enriches understanding and decision-making in financial contexts.

2. **How are stocks in each cluster evaluated relative to each other?** For each cluster generated from the algorithm based on user specified attributes, a variety of stocks from the SP500 database via the yahoo finance API are grouped together. These stocks, when compared on specified attributes, are within the same range. For example, when a cluster of stocks is developed based on 2 specified attributes (i.e Market Cap and Earnings Per Share), each stock in the cluster has a unique market cap and earnings per share with its displayed node size being a function of the relative value of each stock-attribute combination when compared to the averages for the cluster. For stock nodes in the cluster that have very large sizes, they can be investigated as being potentially overvalued relative to the other stocks in the cluster that are smaller in node size.

3. **How well do the stocks grouped in clusters compare to real world returns?** Once a cluster is generated and a group of nodes is displayed, it can then evaluate the stock associated with each node individually to determine its real world return. For example, in the following visualization, where a cluster was generated based upon Revenue, Price to Book, and Operating Margins, the node associated with the company URI is potentially an overvalued investment because it has higher relative values to the overall cluster and thus is larger in size. If we look at the previous month's returns for this stock, as shown in the line chart, we see that it is down 12 percent, potentially indicating that the market is acting accordingly to the hypothesis.

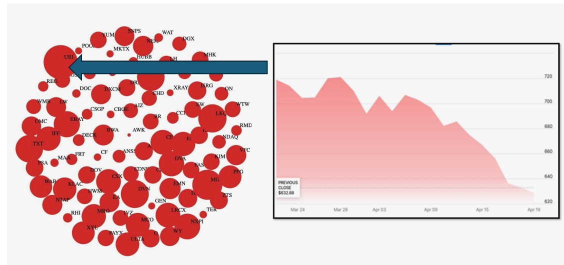

Fig. 5. On the left: clustering output for United Rentals based on revenue, price to book, and operating margin. On the right: stock performance of URI in the past month.

4. **How are similar existing interactive visualizations displayed, and how can we make it more user-friendly using design principles from theirs?** To create a user-friendly interactive visualization tool, it's essential to draw inspiration from similar existing visualizations and incorporate design principles gleaned from them. This involves meticulously studying layout, color schemes, and interactivity features to discern what works well and what could be improved upon. Implementing aesthetically pleasing elements, such as adjusting ticker sizing based on relative averages of cluster attribute values, enhances the tool's visual appeal and aids in pattern recognition. Distinct coloring and intuitive parameter selection UIs further streamline the user experience, allowing users to customize their visualization effortlessly. Conducting thorough user testing, both internally and externally, helps identify areas for improvement, from font choices to interface placement. By iteratively refining the tool based on user feedback, while maintaining a clean and uncluttered design, you can ensure that your interactive visualization effectively communicates information while providing an engaging and intuitive experience for users.

5. **How well do clustered tickers compare to known real-world company connections?**

   Since there is no way to definitively test the "accuracy" of the visualization, a pos-

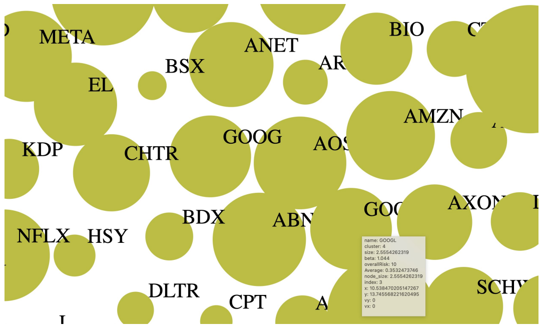

Fig. 6. When clustering is based on beta and overall risk, we see similarities in tech companies' tickers such as META, NFLX, GOOG, GOOGL, and AMZN.

sible test can be randomly sampling clusters and looking at well-known company relationships within each cluster to determine the presence or absence of real world connections. We tested the transitive property of the model by looking at the results of large companies that are proven to be associated along the projected space we're visualizing (e.g. as shown above, Google is next to Amazon as expected, Meta is shown in the same cluster as Netflix). Surface level tests of different clustering settings indicated consistent relationships when clustering tickers, similar to real-world company connections.

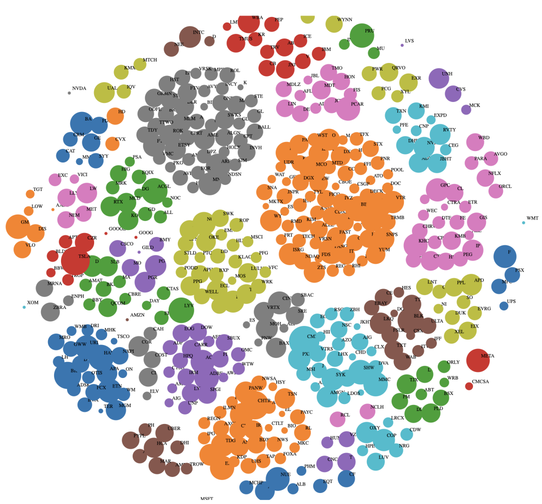

Fig. 7. The visualization with a high number of output clusters selected by the user.

Conclusion and Discussion

A d3 visualisation with a back end that creates an algorithm for S&P 500 stocks, coupled with a drop-down menu for feature selection, represents a sophisticated and powerful tool for investment analysis. The back-end utilizes algorithms and data processing techniques to generate insights and predictions based on user-selected features. Users can interact with the visualization by selecting specific features from the drop-down menu, such as price movements, fundamental ratios, or technical indicators. These features allow the user to tailor their analysis on features which they believe to be the most important, and highlight specific stocks that could be undervalued. Additionally, the project as a whole provides users with valuable insights and predictions based on user selected features. This enables users to make more informed investment decisions, identify potential opportunities, and mitigate risks effectively. When also combined with a summary of sentences by emotion for relevant stocks, both tools go hand in hand to add another dimension of consideration to the investment decisions made. It can also drive further, more targeted and purposeful, research.

Challenges

During this project, I faced a lot of difficulties. Firstly, travelling back home to Saudi Arabia came with a lot of challenges. One future addition I would make is the use of SageMaker for the NLP inference aspect. This was not possible due to location restrictions. The use of AWS would have allowed the use of s3 buckets and a better environment overall to process more data and more 10-K filing.

Secondly, technical difficulties. After developing the visualization aspect, I was unable to connect it to the NLP inference portion and deploy the model due to my laptop breaking down. This meant I had to use my ipad in order to complete the project and could not deploy it or demo the front-end aspect. Even with colab pro and high ram enabled there were issues with the user interface and debugging:

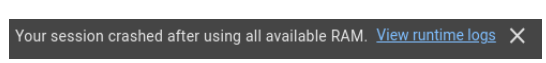

Fig. 8. Crashed google colab