Exploring Investors' Emotions for Financial Sentiment Analysis using FinBERT and Multivariate Time Series Analysis

Harshil Amin – 235848750
Michael Owino – 245808820

Abstract

With the rise and massive commercialization of transformer models, Natural Language Processing has been revolutionized and has changed how we approached NLP tasks in the field. This has further sowed more interests in applying NLP techniques in the finance domain. This paper introduces the use of a finance domain pretrained model, FinBert on a Stock Emotion Dataset which has only been tested and analysed with DistilBert, BERT and RoBERTa recently. Stock Emotions Dataset is a new unique dataset which differs from mostly used dataset on Finance sentiments, meaning it has emojis, fine-grained emotions labels and times series data. We have used FinBERT for financial sentiment and emotion classification and compared to other baselines researchers have used. We have also used a Temporal Attention LSTM Model which combines text, sentiment, emotion and price index for multivariate time series forecasting.

1. Introduction

Finance domain is among the industries that heavily relies on and is influenced by data and stakeholders such as customers and investors which make it a perfect world for using NLP. With NLP financial institutions can handle large amounts of unstructured data to get the latest market trends, predict market shift and sentiment analysis. (Szymanska-Bobowska, 2022).

Some empirical studies in this area of behavioral finance have supported the effects of the social media feelings on the investors and on the market responses. Different authors have highlighted that such views transmitted through this platform contains predictive signals of the future market returns. (Kingstone Nyakurukwa, 2024). Scholars are aiming at forecasting stock prices using results that have been obtained from social media analysis of sentiments and prices of the stocks which have been existing on the market. This has been an issue given the uncertainties involved with the stock market and triggered more desire to determine with precision, the factors that have an impact on the stock market. (Kiersz, 2015)

This comes as a shock and is very hard to predict since, based on the EMH theory, stock markets do not even have a tendency of moving in a straight line. This has forced researchers to look for a model that can predict the correct stock market price. The EMH hypothesis does not explain the continual swings in stock market pricing over short periods since it believes that all information about a stock is factored into the price, and no amount of analysis can help an investor to predict the stock market (Titan, 2015). These movements could be caused by investors purchasing or selling a certain stock in response to news articles or social media reactions. Stock market values are not static but rather dynamic, influenced by news, social media, politics, investor emotions, and the overall economy (Pedro M, 2017). Recent research has shown that online sources such as Twitter can be used to gauge investor sentiment and estimate the price of the Dow Jones Industrial Average (Prodromos, 2016). The amount of views on two Wikipedia financial market pages can be a strong indicator of stock market developments. Investors use these financial market pages to buy and sell equities (Martinez & Cruz, 2022).

The diagram below show's FinBERT architecture

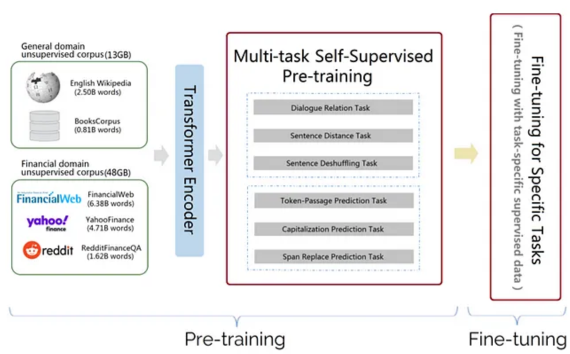

Fig. 1 Pre-training Architecture of FinBERT (Shinde, 2021)

2. Related Work

Studies with regard to the identification and classification of emotion have been explored rather well in the context of NLP, utilizing data from social networks, news, and dialog. Agreeing with the above sources, most of the studies employ annotations of Ekman's 6 basic and Plutchik's 8 basic emotions. A recent work introduces what is currently the largest collection of human-labeled emotions, called GoEmotions (Alon & Ko, 2021)which have been annotated on 27 emotions or none. This work suggests that BERT is better than other comparative solutions, and other works confirm this conclusion. Scellaire social media datasets are as follows : The several existing datasets are informal in language including slang, emoji and hashtags.

Sentiment analysis has become significantly different when extended to the field of finance over the years. Traditional approaches were mostly limited to machine learning approaches such as logistic regression, support vector machines or Naive Bayes (Kearney, 2014). Even though simple, there is always a challenge with these approaches when addressing the complexity and specificity of financial writings. They raised a new paradigm of this discipline with the help of deep learning and transformer models that offer better and more precise evaluations.

In particular, a BERT variation finetuned for financial sentiment analysis is called FinBERT (Araci, 2019). Living up to the challenges posed by the financial text, FinBERT is built, which enhances the performance matrices over the conventional NLP. According to the study conducted by Araci (2019), FinBERT outperforms other state-of-art models in financial communication processing, due to extremely high accuracy and stability of sentiment classification (Araci, 2019).

Inclusion of recursion techniques such as transformer models as in the FinBERT enhances the efficiency and accuracy of market forecasting. This research paper states the perspective of this research and contributes to a relatively limited number of studies urging the application of advanced NLP methods in the sphere of finance to enhance decision-making and strategy development (Socher, Pennington, & Mannings, 2014).

3. Proposed Approach

3.1 Dataset

We have encompassed 2 datasets in our study i.e. StockTwits Dataset and Historical Price data from Yahoo. StockTwits is a social media platform where investors share about opinions stocks and companies. Users of the platform share brief comments that contains sentiment annotations i.e. bullish or bearish and also cash-tags (e.g. $APPL) that helps us identify the associated stock and company. The collected data covers majority of the S&P 500 companies and also has a date attribute that ranges start to end 2020.

The historical price data from Yahoo Finance also encompasses S&P 500 companies in the year 2020. The dataset also contains relevant information such as opening, high, low and closing price of the stocks. Trading volume of the stock is also included in the dataset.

3.2 Data cleaning

We filter out and remove promotional context to the best of our ability. Since the volume of the data was large, we removed rows that had null values on the target class. We also strictly filtered data to the 2020 calendar year. Standardizing timestamps was essential for consistent processing.

3.3 Pre-Processing

3.3.1 Masking

We use special characters for masking, such as a cashtag with a [CTAG] token, a hashtag with a [HTAG], or a website URL with a [URL]. When a comment contains only masked tokens, it is also removed.

3.3.2 Emoji Transformations

Users often express their emotions using emojis in social media and therefore we narrow down the comments consisting of at least one emoji. For instance, in this comment "Google's current form on the market is unpredictable! 😊", it is difficult to detect investor's emotion with text only. Thus, we convert emoji to its textual meaning i.e. " Google's current form on the market is unpredictable! [smiling_face_with_smiling_eyes]. We use the emoji library from python to do emoji transformations.

3.3.3 Tokenization

Tokenization and Length Filtering We limit the sequence length to a maximum of 256 tokens to use FinBERT. Also, we choose comments over 3 tokens long using NLTK's word tokenizer and normalize the token repetitions by having over 4 unique tokens long. This length filtering is applied because short comments normally less convey contextual information.

3.3.4 Emotion Annotations.

We use the 12 emotion classes from the first StockEmotions Dataset (Lee, Youn, Poon, & Han, 2023) who used a multistep pipeline to inspired by human and machine collaboration.

3.4 Data Analysis

3.4.1 Sentiment Distribution After cleaning and filtering our dataset, we remained with 10, 000 rows of data which had around 5,500 bullish sentiments and 4,500 bearish sentiments. This was a good balance meaning our model would learn effectively from both sentiments.

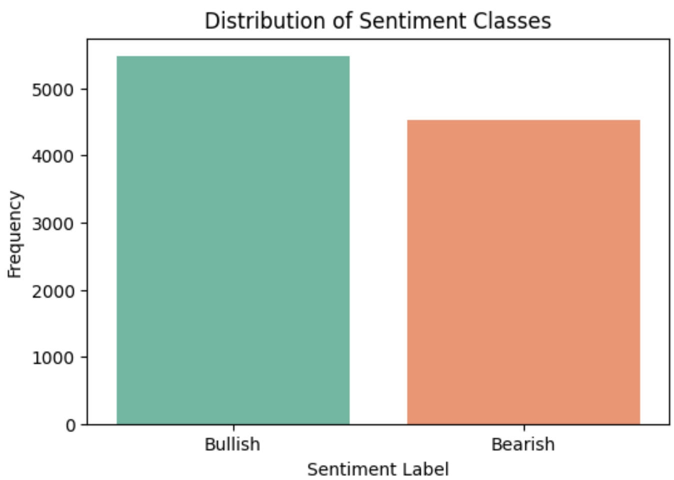

Fig. 3.4.1 Distribution of Sentiment Classes

3.4.2 Emotion Distribution The distribution of emotions identified in the StockTwits dataset show that there is richness in the context of emotions that can be observed in financial social media. Optimism is dominant with the highest frequency of usage, succeeded by anxiety and excitement which may mean the nature of stock market participation is full of hope bearing equal measure of uncertainty. Anger and Disgust are also presented frequently, probably as a result of adverse market experience or a conflict with other members of the community. Where positive and negative extreme emotions are placed tightly interconnecting each other, more scalar negative emotions such as depression or panic are comparatively less observed. This distribution of emotions confirmed that even though, StockTwits users keep positive disposition in trading and investing, they freely post all classes of emotions involved in trading, making the researcher to capture real time trader emotions in social media platforms.

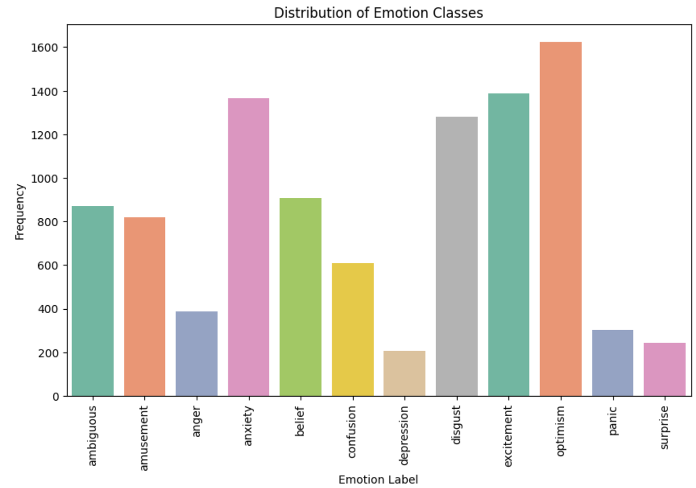

Fig. 3.4.2 Distribution of Emotion Classes

3.4.3 Emotion Correlation The analysis of emotion correlations in the context of the StockTwits is as follows: The heatmap also shows the positive correlated emotional pairs which are ambiguous-amusement, anxiety-confusion with 0.09 and 0.08 respectively. On the other hand, moderate negative correlations exist between the contrasting emotional conditions revealing that low optimism corresponds to high panic, and low anger, to low excitement (-0.28), and (-0.08), respectively. These correlations indicate that some feelings are likely to be found in combination with other feelings when trading is being discussed, while others are quite the opposite. For instance, when users are optimistic, they cannot be panicking at the same time this captures social usual psychological positioning as it pertains to investment topics.

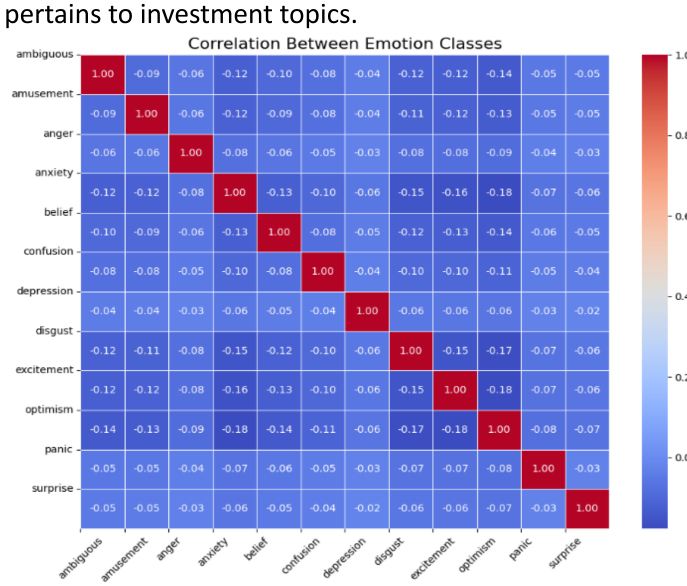

Fig. 3.4.3 Correlation Matrix for Emotion Classes

3.4.4 Sentiment Distribution Different Companies Study found that distribution of tweets nowhere near balanced as trending companies had many tweets like (TSLA) Tesla while less trending companies like PG(Procter & Gamble Co) had very few tweets as show in the figure below.

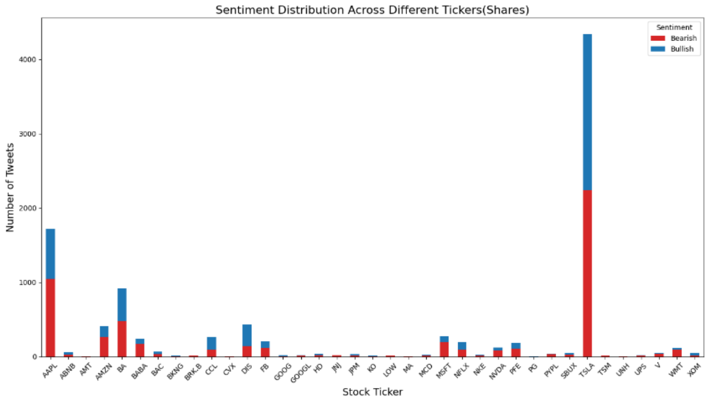

Fig. 3.4.4 Sentiment Distribution across different Tickers

4. Modelling

4.1 Financial Sentiment Analysis

This task introduces a novel multi-task learning approach utilizing FinBERT, a language model specifically designed for financial domain text analysis. By developing a custom neural network architecture with dual classification heads, the research aims to simultaneously predict both sentiment and emotional labels from financial text. The proposed model leverages the contextual understanding inherent in FinBERT, enabling a more nuanced interpretation of financial communication beyond traditional single-task sentiment analysis.

The methodology employs a comprehensive tokenization process using the FinBERT tokenizer, with texts processed to a maximum sequence length of 256 tokens. The model's architecture strategically extracts features from the [CLS] token's hidden representations, routing these through task-specific linear classifiers dedicated to sentiment and emotion prediction. This approach allows for a more sophisticated analysis that captures the complex linguistic subtleties present in financial discourse, moving beyond simplistic binary sentiment classifications.

The training methodology incorporates a multi-task learning strategy, utilizing a combined cross-entropy loss function that enables simultaneous optimization of sentiment and emotion prediction tasks. Hyperparameters were tuned numerously, including learning rates, weight decay, and training epochs, to enhance model performance and generalizability. By implementing a nuanced approach that considers both emotional tone and sentiment, this research provides a more comprehensive framework for understanding the linguistic characteristics of financial text, potentially offering more refined insights for financial communication analysis.

4.2 Multivariate time Series

This task explored multivariate time series analysis approach for financial forecasting, integrating advanced deep learning techniques with temporal sentiment analysis. The methodology employs a comprehensive data preprocessing pipeline that combines stock price data from multiple sources and incorporates sentiment and emotional labels. By utilizing a sliding window approach, the research develops a novel method for aggregating and normalizing financial time series data, transforming raw financial indicators into a structured, machine-learning-ready format.

We have used a Temporal Attention LSTM (Long Short-Term Memory) model, which introduces an mechanism to capture complex temporal dependencies in financial time series. The model architecture is designed to dynamically weight different time steps, allowing for more nuanced feature extraction from historical stock price movements and sentiment indicators. By incorporating bidirectional LSTM layers and a custom attention mechanism, the approach aims to overcome traditional limitations in time series prediction by providing a more context-aware predictive framework.

The training methodology implements advanced techniques such as gradient clipping and mini-batch gradient descent to enhance model stability and generalization. The model is configured to process a multivariate input space of 40 features, including normalized stock prices, sentiment labels, and emotional indicators, with a sophisticated loss function based on Mean Squared Error.

5. Experimental Analysis

5.1 Sentiment Analysis

For the classification task of emotion and sentiment, we randomly split the data into train/validation/test sets in the proportions of 80%/10%/10%. The total data containes 10,000 rows after cleaning and pre-processing.

Our binary sentiment classification model demonstrates strong performance with an **accuracy of 75%** across the validation set, showing balanced precision and recall scores for both positive and negative sentiments (f1-scores of 0.77 and 0.71 respectively). The confusion matrices reveal that the model performs slightly better at identifying negative sentiments **(428 true negatives)** compared to positive sentiments **(318 true positives),** with relatively balanced misclassifications.

In contrast, the emotion classification task proves more challenging, achieving an overall **accuracy of 38%** across twelve emotion categories. The model shows varying performance across different emotions, with the best f1-scores seen in emotions like confusion (0.49) and anxiety (0.49), while struggling with emotions like surprise (0.07) and amusement (0.22). This performance disparity between sentiment and emotion classification suggests that while the model can effectively distinguish between positive and negative sentiments, capturing the nuanced differences between specific emotions remains a more complex challenge.

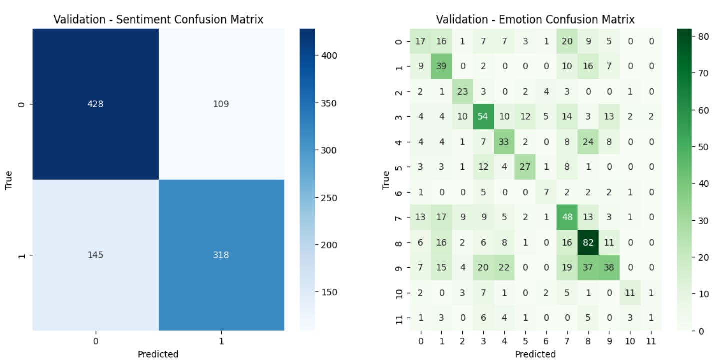

Fig. 5.1 Validation of (i) Sentiment Confusion Matrix, (ii) Emotion Confusion Matrix

Due to the imbalanced nature of the data, we observed that less frequent emotions, such as ambiguity and surprise, are more likely to be confused with one another.

5.2 Multivariate Time Series

We implemented a Temporal Attention LSTM doing the Multivariate Analysis Task. It did exceptionally well and was able to surpass the benchmarks.

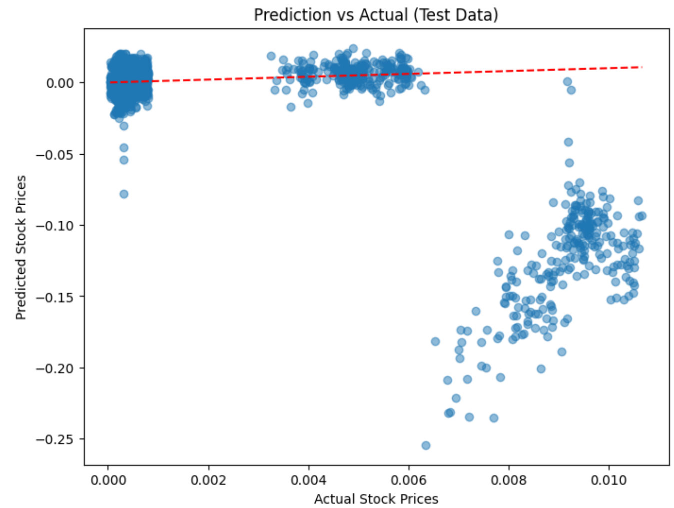

Fig. 5.2.1 Prediction vs Actual (Test Data)

The prediction vs actual plot reveals above interesting patterns in the model's performance on test data, with predictions clustering in three main regions. While the model shows good accuracy for lower stock prices (clustering around 0.000-0.006), it tends to underestimate values in the higher price range (0.008-0.010), with predictions falling significantly below the actual values (shown by the deviation from the red dashed line). The overall performance metrics are encouraging, with a low Mean Squared Error of 0.0014 and Mean Absolute Error of 0.0134, indicating that the model's predictions, while not perfect, maintain reasonable accuracy in capturing price movements based on the combined sentiment, emotion, and historical price signals.

The learning curve diagram below demonstrates consistent model improvement over 200 epochs, with the loss steadily decreasing from approximately 0.00679 to 0.00665. This smooth, downward trajectory suggests that the Temporal Attention LSTM effectively learned the temporal patterns in the merged sentiment, emotion, and price data without signs of overfitting, as there is no evident plateau or increase in the loss value.

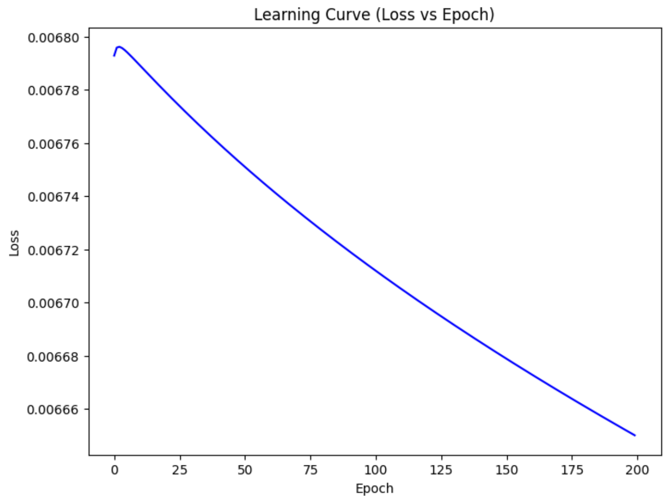

Fig. 5.2.2 Learning Curve (Loss vs Epoch)

This performance beats the benchmark (Lee, Youn, Poon, & Han, 2023) **which had a Mean Squared Error of 0.83 while ours was 0.0014 meaning our LSTM model better.**

6. Conclusion

The FinBERT model performed really well and almost nearing the score the pretrained models benchmarks given the short training time and less fine tuning time. This comes impressive given the fact that the nature of dataset caused the model to have limitations. FinBERT is originally trained on three labels (bullish, bearish and neutral) but our dataset had only bullish and bearish. The dataset impacted models' predictive capability negatively, reduced granularity in sentiment classification and decreased the overall performance and generalizability.

The temporal attention LSTM model did very well beyond our expectations. Temporal Fusion Transformers are proved to be the best and more powerful in handling time-series analysis tasks. Exploration of Temporal Fusion Transformers remain for future research.

7. References

[1]. Araci, D. (2019). *FinBERT: A Pretrained Language Model for Financial Communications*.

[2]. Demszky, D., et al. (2020). *GoEmotions: A Dataset for Fine-Grained Emotion Classification*.

[3]. Szymanska-Bobowska, A. (2022). *NLP Applications in Finance*.

[4]. Bezerra, A., & Albuquerque, P. (2017). *Stock Market Dynamics*.

[5]. Thune, N. (2020). *Efficient Market Hypothesis and Stock Price Predictions*.

[6]. Chourmouziadis, K., & Chatzoglou, P. (2016). *Social Media Sentiment and Stock Price Estimations*.

[7]. Hao, J. (2020). *Wikipedia Page Views as Market Indicators*.

[8]. Kiersz, A. (2015). *Volatility and Stock Market Influences*.

[9]. Dataset Reference:

- Kaggle Dataset. (n.d.). Stock Emotions Dataset. Retrieved from https://www.kaggle.com/datasets/goddiao/stock emotions