Context-Specific Sentiment Analysis of Financial Data

Prayash Joshi
Department of Statistics
Virginia Tech
prayash@vt.edu

Abstract

In this study, we explores the development of a model capable of conducting context-specific sentiment analysis on financial news to assess market sentiments more accurately. This is done using a unified model that simultaneously predicts a topic and sentiment labels are predicted. Research has shown that Financial markets are can be influenced by news articles and public sentiments. Traditional sentiment analysis approaches often overlook the intricate relationships between news topics and sentiment, leading to potential misinterpretations in sentiment assessment. Our research addresses this gap by proposing a unified model that integrates sentiment and topic classification in a single framework, aiming to enhance the predictive power and relevance of sentiment analysis in financial contexts. The model leverages a deep learning architecture combining LSTM and CNN layers to process and analyze textual data from varied financial news sources. We explore the model's effectiveness through extensive validation on multiple datasets, highlighting its capabilities and limitations in handling real-world financial texts. Using our smaller dataset of AAPL news, we've constructed a intuitive interacative dashboard for users to understand the impacts of sentiment analysis.

1 Introduction

There is a clear impact of news on financial markets that presents a complex but critical challenge in financial analysis. In the real world, news articles can swiftly influence investor behavior and market trends. Thus, traders and investors alike are interested in a sentiment analysis type of tool. Traditional sentiment analysis primarily focuses on determining the overall sentiment of texts—categorizing them into positive, negative, or neutral sentiments. However, this approach often fails to capture the context-specific nuances that are critical in financial domains, where the relevance of news topics can significantly influence the sentiment interpretation.

The motivation for this study stems from the need to enhance sentiment analysis techniques by incorporating context-aware methodologies that can differentiate and weigh the sentiments based on the associated topics. For instance, news about an 'earnings increase' for a company holds different implications and weight compared to 'technological advancements' within the same company, necessitating a model that can understand and integrate these nuances.

Our research contributes to this area by developing a unified model that not only predicts sentiment but also classifies the topics of financial news articles, allowing for a more nuanced analysis. This approach helps in better understanding the interplay between news topics and their impact on market sentiment. We employ a hybrid deep learning architecture that integrates LSTM (Long Short-Term Memory) networks for sequence processing with CNNs (Convolutional Neural Networks) for feature extraction, addressing both the sequential and spatial aspects of textual data.

STAT4984 , LateX template borrowed from NIPS 2017.

Our study also develops a dashboard visualization that is publically avaliable for all users based on our smaller dataset of AAPL news.

The following sections detail the datasets used, including their composition and preprocessing steps, the methodology behind our model's construction, the results of our experiments compared to traditional models, and a discussion on the implications of our findings and potential future research directions.

2 Related work

Sentiment analysis, a core component of natural language processing, focuses on extracting and interpreting emotions or attitudes conveyed through text. This technology plays a crucial role across various domains such as marketing, public opinion analysis, and particularly in the financial sector, where it aids in stock market predictions, investment decision-making, and analyzing customer opinions on financial products [3]. Despite its vast applications, the accurate analysis of financial texts is challenging due to the complex nature of language used, the presence of domain-specific jargon, and subtle expressions of sentiments [4].

The evolution of sentiment analysis techniques has moved from traditional machine learning methods, such as Support Vector Machines (SVM) and Naive Bayes, which depend heavily on hand-crafted features, to more sophisticated deep learning approaches that minimize the need for manual feature engineering. Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks are at the forefront of this shift. CNNs effectively learn relevant features from text data through convolutional filters that capture contextual information and identify key sentiment indicators. Meanwhile, LSTMs excel in capturing long-term dependencies within text, which are crucial for understanding sentiments in financial news [2, 4]. Moreover, hybrid models that combine CNNs and LSTMs leverage the strengths of both architectures to enhance the accuracy of sentiment classification tasks [4].

However, the field faces significant challenges, particularly in the financial domain where texts often contain complex sentence structures and implicit sentiments. The accuracy of sentiment analysis in this domain can be inconsistent, with models' performance varying widely based on dataset characteristics such as text length, sentiment class distribution, and annotation quality [2]. Additionally, adapting these models to new financial domains often requires extensive efforts in data annotation and model tuning [1].

Looking ahead, the development of multi-output models that can predict both sentiment polarity and intensity simultaneously is a promising direction for providing more nuanced insights into financial texts [1]. Furthermore, integrating domain-specific features such as financial performance indicators with textual features has been shown to improve classification accuracy [3]. This integration is particularly relevant to your research, where enhancing a CNN-LSTM architecture with such features could yield substantial benefits. Additionally, leveraging transfer learning techniques to utilize pre-trained models like BERT and FinBERT may enhance the generalization capabilities of sentiment analysis models within the financial domain [4]. Finally, evaluating these advanced models across diverse financial datasets and comparing them with state-of-the-art benchmarks will be crucial in demonstrating their effectiveness and applicability in real-world scenarios.

3 Dataset and Features

The datasets employed in this study comprise financial news articles and sentiment labels, meticulously curated from diverse sources. Our primary dataset, designated as `combined_data`, was sourced using the AlphaVantage API. It initially featured a variety of fields such as `title`, `time_published`, `summary`, `source`, `relevance_score`, `ticker_sentiment_score`, `ticker_sentiment_label`, `topic`, and `ticker`. Following a thorough cleaning process that involved removing less significant columns and extracting the primary topic for each article, the dataset was refined to include `title`, `summary`, `topics`, `ticker_sentiment_score`, `ticker_sentiment_label`, and `mapped_label`, resulting in a total of 1,957 entries.

Additionally, the `all_phrases` dataset was incorporated to enrich the data pool. Originating from a renowned financial analysis platform, this dataset required conversion from text format to a structured

dataframe, featuring `summary`, `sentiment`, and a newly introduced `confidence` column derived from agreement percentages noted in the filename, comprising 17,362 rows.

To further augment the data, additional records were retrieved for 19 different stock tickers across a variety of topics via the AlphaVantage API. The preprocessing steps mirrored those applied to the `combined_data`, ensuring consistency across datasets. These records were combined with AAPL news from the `combined_data` to form the `final_dataset`, which encapsulates a comprehensive array of topics such as Blockchain, Earnings, and Technology, totaling 14,892 rows.

The preprocessing regimen applied to these datasets was extensive and included:

- Parsing JSON formatted data from API calls,
- Synchronizing datasets to align with the financial phrases data for our initial model,
- Developing a bespoke algorithm to accurately determine the primary topic from a list in the `final_dataset`,
- Encoding sentiment and topic labels as integers and subsequently converting these into categorical format for compatibility with the `categorical_crossentropy` loss function,
- Comprehensive text preprocessing that involved tokenization, case normalization, punctuation removal, stopword filtering, lemmatization, and the cleansing of HTML tags and special characters from summaries,
- Imputing missing values to maintain data integrity,
- Harmonizing label formats from the API with those used in the financial phrases dataset.

For model training and evaluation, the data was partitioned into training and testing sets following an 80/20 split, applied to the three models developed in this study.

4 Methods

4.1 Model 1: Sentiment Analysis Model on Small Financial Data + Financial Phrases

Model 1 is designed to classify the sentiment of financial texts using a deep learning architecture that leverages a combination of convolutional and recurrent neural networks. The primary objective is to understand and accurately classify the sentiment expressed in financial news and phrases into predefined categories: positive, neutral, and negative.

Data Collection and Preparation The dataset for this model comprises two main sources: a collection of financial phrases with pre-labeled sentiment and real-time financial news data fetched using the AlphaVantage API. Initially, the data is gathered in raw format, which includes various metadata fields. Preprocessing steps involve cleaning the data by removing unnecessary metadata, normalizing the text (removing punctuation, converting to lowercase, etc.), and consolidating all summaries into a single text corpus.

Text Processing The processed text data undergoes tokenization, where texts are converted into sequences of integers. Each integer represents a unique word in a dictionary formed from the entire text corpus. These sequences are then padded to ensure uniformity in sequence length, a necessary step for training neural networks.

Model Architecture The model architecture comprises the following layers:

Table 1: Neural Network Architecture of Model *basic_sentiment_model*

Layer (type)	Output Shape	Param #	Connected to
input_layer (InputLayer)	(None, 100)	0	-
embedding_1 (Embedding)	(None, 100, 100)	variable	input_layer[0][0]
conv1d_1 (Conv1D)	(None, 96, 64)	variable	embedding_1[0][0]
maxpooling1d_1 (MaxPooling1D)	(None, 24, 64)	0	conv1d_1[0][0]
lstm_1 (LSTM)	(None, 24, 50)	variable	maxpooling1d_1[0][0]
dropout_1 (Dropout)	(None, 24, 50)	0	lstm_1[0][0]
lstm_2 (LSTM)	(None, 50)	variable	dropout_1[0][0]
dense_1 (Dense)	(None, num_categories)	variable	lstm_2[0][0]

Compilation and Training The model is compiled using the Adam optimizer and categorical crossentropy as the loss function, reflecting the multi-class nature of the sentiment labels. Metrics such as accuracy, precision, and recall are monitored. Training involves an Early Stopping mechanism to prevent overfitting, with a patience parameter set to 5 epochs. Additionally, the best model configuration is saved using Model Checkpoint based on validation accuracy.

Performance Evaluation Upon training, the model's performance is evaluated on a held-out test set to assess its ability to generalize beyond the training data. The evaluation metrics used include loss, accuracy, precision, and recall, providing a holistic view of model effectiveness.

Summary: This initial model serves as a foundational step in understanding and categorizing sentiments in financial texts, aiming to enhance the responsiveness and accuracy of sentiment-based financial analysis tools.

4.2 Model 2: Unified Neural Network Model on Small Financial Dataset

Model 2 was developed to perform both topic classification and sentiment analysis simultaneously, utilizing a unified neural network architecture. This model was trained on a small dataset consisting of approximately 2000 financial news articles related to Apple Inc. (AAPL), each tagged with sentiment labels and topics.

Data Preparation: The dataset comprised summaries of financial news, which were preprocessed before training. Text preprocessing involved removing HTML tags, special characters, and converting text to lowercase to standardize the input data. Furthermore, the NLTK library was used for tokenization, and words were filtered through a list of stopwords to remove uninformative words. Each word was then lemmatized to reduce it to its base or dictionary form. The prepared texts were then tokenized using Keras's text processing utilities, and sequences were padded to ensure uniform input size.

Model Architecture:

Table 2: Neural Network Architecture of Model *functional_17*

Layer (type)	Output Shape	Param #	Connected to
input_layer_2 (InputLayer)	(None, 100)	0	-
embedding_4 (Embedding)	(None, 100, 100)	538,500	input_layer_2[0][0]
bidirectional (Bidirectional)	(None, 128)	84,480	embedding_4[0][0]
dropout_4 (Dropout)	(None, 128)	0	bidirectional[0][0]
dense_4 (Dense)	(None, 128)	16,512	dropout_4[0][0]
dense_5 (Dense)	(None, 128)	16,512	dropout_4[0][0]
topic_output (Dense)	(None, 3)	387	dense_4[0][0]
sentiment_output (Dense)	(None, 3)	387	dense_5[0][0]

4.3 Model 3: Modified-Unified Model on Large Financial Dataset

Model 3 extends the unified approach of Model 2 by incorporating a broader dataset with enhanced preprocessing and a simplified model architecture to improve generalization across a more diverse set of financial news topics.

Data Collection and Expansion: To enrich the training data and improve model robustness, additional data was collected using the AlphaVantage API, covering a wide range of financial sectors and companies. The data included sentiment and topical information for companies like Microsoft, Google, Amazon, and smaller tech startups, spanning from 2019 to the present year. This expansion aimed to provide a comprehensive view of market sentiment across different industry segments.

Data Preprocessing: The raw data from the API was initially in JSON format, containing various metadata along with the main content. The preprocessing steps involved:

- Extracting relevant fields such as news summary, sentiment scores, and topic relevance.
- Normalizing text by converting to lowercase, removing punctuation, and applying the NLTK library for tokenization and lemmatization.
- Mapping sentiment labels to a uniform set of categories (positive, neutral, negative) using predefined mappings.
- Simplifying topic data by selecting the primary topic based on relevance scores, ensuring diversity beyond dominant categories through a controlled random selection process.

Model Architecture: The architecture for Model 3 was designed to be less complex than its predecessors while maintaining effectiveness. It includes:

Table 3: Neural Network Architecture of Model *functional_18*

Layer (type)	Output Shape	Param #	Connected to
input_layer_15 (InputLayer)	(None, 100)	0	-
embedding_17 (Embedding)	(None, 100, 100)	1,680,800	input_layer_15[0][0]
bidirectional_13 (Bidirectional)	(None, 64)	34,048	embedding_17[0][0]
dropout_18 (Dropout)	(None, 64)	0	bidirectional_13[0][0]
dense_29 (Dense)	(None, 64)	4,160	dropout_18[0][0]
dense_30 (Dense)	(None, 64)	4,160	dropout_18[0][0]
topic_output (Dense)	(None, 14)	910	dense_29[0][0]
sentiment_output (Dense)	(None, 3)	195	dense_30[0][0]

Training and Optimization: The training process involved:

- Using the Adam optimizer with an exponential decay schedule to adjust the learning rate over epochs, helping in stabilizing the model's convergence.
- Applying a 50% dropout rate to reduce overfitting.
- Employing early stopping to halt training if the validation accuracy did not improve for consecutive epochs, ensuring that the model does not overfit to the training data.

Model training was monitored using accuracy, precision, and recall metrics for both outputs, with performance validation conducted on a separate test set to ensure the model's generalizability.

Model 3 represents a sophisticated approach to sentiment and topic classification in financial texts, aiming to deliver robust performance across diverse data sets. By simplifying the model architecture and incorporating a broader data set, it addresses the limitations of earlier models and sets the stage for more nuanced analyses of financial sentiment and topics.

5 Results

5.1 Model 1: Sentiment Analysis Model on Small Financial Data + Financial Phrases

The first model, developed to classify sentiment in financial news texts, exhibited strong performance across multiple metrics. We utilized a combination of convolutional and recurrent layers to capture

both local features and sequence dependency in the data. The training process involved monitoring both training and validation metrics, where early stopping was employed to prevent overfitting.

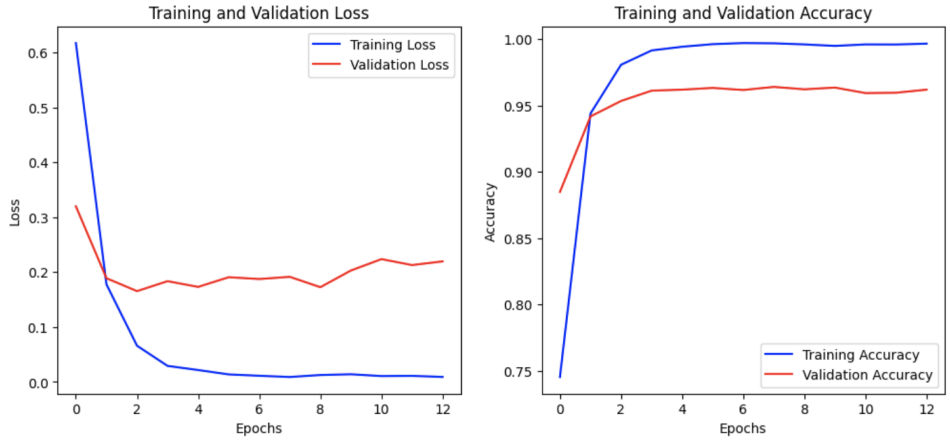

Figure 1: Training and validation loss and accuracy over epochs for Model 1.

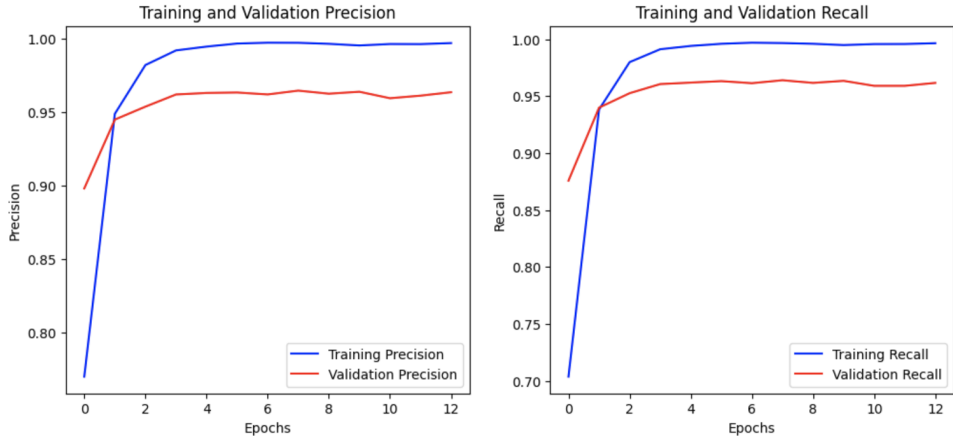

Figure 2: Training and validation precision and recall over epochs for Model 1.

Despite achieving high training accuracy, the validation results showed a slight lag, indicating a minor overfit to the training data. This was addressed by introducing dropout layers which helped in regularizing the model. Precision, recall, and F1-scores are depicted below in a structured tabular format, highlighting the model's ability to generalize across different sentiment classes.

Table 4: Classification Report for Model 1

Class	Precision	Recall	F1-Score	Support
Negative	0.98	0.95	0.97	453
Neutral	0.97	0.98	0.97	2315
Positive	0.95	0.94	0.94	1096
Accuracy		0.96		
Macro Avg	0.97	0.96	0.96	3864
Weighted Avg	0.96	0.96	0.96	3864

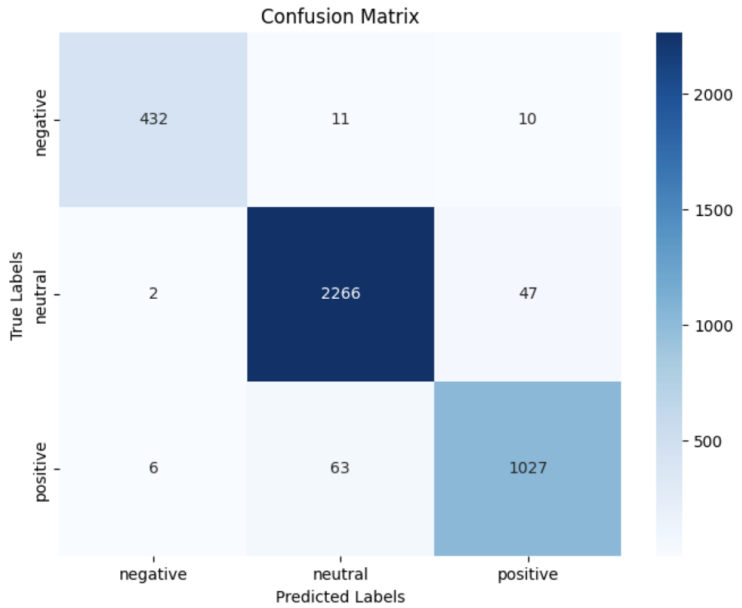

Figure 3: Confusion matrix for Model 1, illustrating class-wise performance.

Further validation through 5-fold cross-validation affirmed the model's robustness, with an average accuracy of 95.91% and minimal variance, demonstrating the model's effectiveness and stability across different subsets of the dataset.

Table 5: 5-Fold Cross-Validation Results for Model 1

Fold	Accuracy (%)	Precision (%)	Recall (%)
1	96.27	96.30	96.22
2	95.78	95.81	95.76
3	96.14	96.16	96.04
4	95.16	95.16	95.13
5	96.19	96.22	96.19
Mean	95.91	95.93	95.87

These results demonstrate the capability of the model to effectively analyze sentiments within financial news articles, highlighting its potential utility in financial analytics and decision-making processes.

5.2 Model 2 and 3: Unified Models:

Table 6: Comparison of Accuracy in Training vs. Validation for Sentiment and Topic

Model	**Testing**		**Validation**	
	Sentiment	Topic	Sentiment	Topic
Model 2	0.9543	0.7416	0.7066	0.4260
Model 3	0.9309	0.9212	0.7130	0.5519

Model 3 outperforms Model 2 in terms of overall balance and robustness, particularly in handling a wider array of topics. The improved topic classification accuracy in Model 3 suggests that enhancements in data preprocessing and model training strategies have effectively addressed some of the overfitting issues observed in Model 2.

6 Conclusion and discussion

Insights and Model Performance: Our initial model in excellent in understanding financial phrases and jargon. It can predict the correct sentiment label 95 percent of the time. This sets us up nicely for future work where we can integrate an ensemble of neural networks that can help learn specific things. But, due to the complexity of this ensemble approach, I decided to build a unified neural network that predicts the topic and sentiment label simultaneously instead. This model had a slightly lower prediction accuracy for sentiment labels but introduced topics can be instrumental for analysis.

Advancements in Model Techniques: Model 3's advancement over Model 2, particularly in handling a larger and more varied dataset, illustrates the importance of dataset quality over sheer quantity. Moreover, the iterative refinement of the models highlighted the trade-offs between model complexity and performance, where simpler, well-tuned models often surpassed more complex configurations.

Future Research Directions: Future research will focus on using the information gained from model 1 to improve the sentiment results and topic understandings in model 2 and 3. Additionally, integrating more dynamic data elements such as real-time market sentiment, global economic indicators, and cross-asset influences to enhance the models' predictive accuracy and robustness. Finally, I would like to improve the dashboard to have more stocks.

7 Code Availability

The source code and additional resources used in this study are available on GitHub below:

https://github.com/PrayashJoshi/Context-Specific-Sentiment-Analysis-in-Financial-News

8 Dashboard Availability

The following website guidelines the isights gained from the second model to map sentiment analysis along with AAPL news

https://stocknews.pages.dev/

References

[1] Taqwa Hariguna and Athapol Ruangkanjanases. Adaptive sentiment analysis using multioutput classification: A performance comparison. *PeerJ Computer Science*, 9:e1378, 2023.

[2] Hannah Kim and Young-Seob Jeong. Sentiment classification using convolutional neural networks. *Applied Sciences (Switzerland)*, 9, Jun 2019.

[3] Srikumar Krishnamoorthy. Sentiment analysis of financial news articles using performance indicators. *Knowledge and Information Systems*, 56(2):373–394, Aug 2018.

[4] Ishaani Priyadarshini and Chase Cotton. A novel lstm–cnn–grid search-based deep neural network for sentiment analysis. *The Journal of Supercomputing*, 77(12):13911–13932, Dec 2021.