CSE 464 - ADVANCED DEEP LEARNING

Sentiment Analysis

Financial News

Done By: Aadithyaa Prabha (E0321008) & Deevna Reddy (E0321053)

Try Pitch

Try Pitch

INTRODUCTION

Sentiment analysis has become an essential tool in the financial industry, where understanding the sentiment behind news articles, reports, and social media posts can significantly impact trading decisions and investment strategies. The ability to predict market trends and movements based on the sentiment expressed in financial news allows investors, traders, and analysts to make more informed and timely decisions. In this project, we focus on analyzing the sentiment of financial news articles using state-of-the-art natural language processing (NLP) models. Our objective is to classify these articles into sentiment categories such as positive, negative, and neutral. By leveraging advanced models like FinBERT and RoBERTa, we aim to provide a detailed and accurate sentiment analysis that can serve as a valuable resource for market participants. This analysis not only helps in understanding market sentiment but also aids in forecasting market trends, ultimately contributing to more strategic decision-making in the financial sector.

Try Pitch

LITERATURE REVIEW

Traditional methods of sentiment analysis often rely on lexicon-based approaches, where predefined dictionaries categorize words as positive, negative, or neutral. For instance, the use of domain-specific dictionaries has been shown to enhance the accuracy of sentiment classification in financial contexts. However, these methods can struggle with the nuanced language of financial news, which may lead to misclassification of sentiment due to the complex interplay of words and phrases.

Recent studies have introduced more sophisticated techniques, including machine learning and deep learning models. For example, Srikumar Krishnamoorthy et al. developed a hierarchical sentiment classifier that combines dictionary-based methods with machine learning to improve sentiment prediction accuracy . Additionally, the use of performance indicators in conjunction with sentiment analysis has been proposed to provide a more comprehensive understanding of market sentiment

Try Pitch

DESCRIPTION OF DATASET

The dataset used for this project consists of financial news articles, which are labeled with sentiment scores. These articles are collected from various financial news sources, including reputable websites and financial blogs. The CSV file contains two columns with 4845 entries each. The dataset is divided into three sentiment categories: positive, negative, and neutral.

- **Description**: This column contains sentiment labels for the text data. It indicates the sentiment of the text in the corresponding row.
- **Values**: The values in this column can be "neutral", "positive", or "negative".
- **Data Type**: Object (string)

Try Pitch

DESCRIPTION OF DATASET

Sample Rows of the dataset:

neutral	According to Gran, the company has no plans to move all production to Russia, although that is where the company is growing.
neutral	Technopolis plans to develop in stages an area...
negative	The international electronic industry company ...
positive	With the new production plant the company woul...
positive	According to the company 's updated strategy f...
positive	FINANCING OF ASPOCOMP 'S GROWTH Aspocomp is ag...

Try Pitch

TRANSFORMERS

Attention Mechanism for Context Capture:

- Utilizes self-attention to understand word relationships within a sentence.
- Captures long-range dependencies, improving comprehension of context

Scalability and Improved NLP Task Performance:

- Handles large-scale data efficiently, enhancing model performance.
- Outperforms previous models (RNNs, LSTMs) in tasks like sentiment analysis, translation.

Hugging Face Transformers Library

- Offers a wide range of pre-trained models and tokenizers for various NLP tasks.
- Supports seamless integration with popular ML frameworks (PyTorch, TensorFlow).

Try Pitch

METHODOLOGY & APPROACH

1. Data Preprocessing
   a. **Cleaning & Standardization of Text:**
      - Remove noise: punctuation, stop words, special characters.
      - Convert text to lowercase to ensure uniformity.

   b. **Tokenization for Model Input:**
      - Split text into tokens (words/subwords) suitable for model input.

2. Model Training
   a. **Hyperparameter Tuning and Training for Both Models**:
      - Optimize hyperparameters: learning rate, batch size, epochs.
      - Train FinBERT and RoBERTa models on the preprocessed dataset.

Try Pitch

METHODOLOGY & APPROACH

3. Model Evaluation
    a. **Confusion Matrix and Classification Report**:
        - Evaluate model performance using confusion matrix to visualize predictions.
        - Generate classification report detailing precision, recall, and F1-score.

Try Pitch

RESULTS

(a) FinBERT Results:

- Accuracy: FinBERT achieved an accuracy of 0.85 on the validation set, indicating that it correctly classified 85% of the news articles.

- Precision: The precision scores for the positive, negative, and neutral classes were 0.84, 0.87, and 0.83, respectively. This indicates that FinBERT was able to accurately identify positive, negative, and neutral sentiments.

- Recall: The recall scores for the positive, negative, and neutral classes were 0.83, 0.88, and 0.84, respectively. This indicates that FinBERT was able to correctly identify most of the positive, negative, and neutral sentiments in the news articles.

- F1-Score: The F1-scores for the positive, negative, and neutral classes were 0.84, 0.87, and 0.83, respectively. This indicates a good balance between precision and recall for each class.

Try Pitch

RESULTS

(a) FinBERT Results:

- Accuracy: FinBERT achieved an accuracy of 0.85 on the validation set, indicating that it correctly classified 85% of the news articles.

- Precision: The precision scores for the positive, negative, and neutral classes were 0.84, 0.87, and 0.83, respectively. This indicates that FinBERT was able to accurately identify positive, negative, and neutral sentiments.

- Recall: The recall scores for the positive, negative, and neutral classes were 0.83, 0.88, and 0.84, respectively. This indicates that FinBERT was able to correctly identify most of the positive, negative, and neutral sentiments in the news articles.

- F1-Score: The F1-scores for the positive, negative, and neutral classes were 0.84, 0.87, and 0.83, respectively. This indicates a good balance between precision and recall for each class.

Try Pitch

RESULTS

(b) RoBERTa Results:

- Accuracy: RoBERTa achieved an accuracy of 0.82 on the validation set, indicating that it correctly classified 82% of the news articles.

- Precision: The precision scores for the positive, negative, and neutral classes were 0.81, 0.84, and 0.80, respectively. This indicates that RoBERTa was able to accurately identify positive, negative, and neutral sentiments.

- Recall: The recall scores for the positive, negative, and neutral classes were 0.80, 0.85, and 0.81, respectively. This indicates that RoBERTa was able to correctly identify most of the positive, negative, and neutral sentiments in the news articles.

- F1-Score: The F1-scores for the positive, negative, and neutral classes were 0.81, 0.84, and 0.80, respectively. This indicates a good balance between precision and recall for each class.

Try Pitch

OUTPUT

```
predicted_sentiments = []
for text in df['Text']  Loading...
    score = sia.polari     (text)
    if score['compound'] >= 0.05:
        predicted_sentiments.append('positive')
    elif score['compound'] <= -0.05:
        predicted_sentiments.append('negative')
    else:
        predicted_sentiments.append('neutral')

df['predicted_sia'] = predicted_sentiments
```

	Sentiment	Text	Cleaned_Text	Tokenized_Text	Stemmed_Text	Lemmatized_Text	label	predicted_sia
0	neutral	According to Gran , the company has no plans t...	according to gran the company has no plans to...	[according, gran, company, plans, move, produc...	[accord, gran, compani, plan, move, product, r...	[according, gran, company, plan, move, product...	1	negative
1	neutral	Technopolis plans to develop in stages an area...	technopolis plans to develop in stages an area...	[technopolis, plans, develop, stages, area, le...	[technopoli, plan, develop, stage, area, less,...	[technopolis, plan, develop, stage, area, le, ...	1	negative
2	negative	The international electronic industry company ...	the international electronic industry company ...	[international, electronic, industry, company,...	[intern, electron, industri, compani, elcoteq,...	[international, electronic, industry, company,...	2	neutral
3	positive	With the new production plant the company woul...	with the new production plant the company woul...	[new, production, plant, company, would, incre...	[new, product, plant, compani, would, increas,...	[new, production, plant, company, would, incre...	0	positive
4	positive	According to the company 's updated strategy f...	according to the company s updated strategy fo...	[according, company, updated, strategy, years,...	[accord, compani, updat, strategi, year, baswa...	[according, company, updated, strategy, year, ...	0	positive
...	...	...	...	...	...	...	...	...
4841	negative	LONDON MarketWatch -- Share prices ended lower...	london marketwatch share prices ended lower i...	[london, marketwatch, share, prices, ended, lo...	[london, marketwatch, share, price, end, lower...	[london, marketwatch, share, price, ended, low...	2	negative
4842	neutral	Rinkuskiai 's beer sales fell by 6.5 per cent ...	rinkuskiai s beer sales fell by per cent to ...	[rinkuskiai, beer, sales, fell, per, cent, mil...	[rinkuskiai, beer, sale, fell, per, cent, mill...	[rinkuskiai, beer, sale, fell, per, cent, mill...	1	neutral
4843	negative	Operating profit fell to EUR 35.4 mn from EUR ...	operating profit fell to eur mn from eur mn ...	[operating, profit, fell, eur, mn, eur, mn, in...	[oper, profit, fell, eur, mn, eur, mn, includ,...	[operating, profit, fell, eur, mn, eur, mn, in...	2	positive
4844	negative	Net sales of the Paper segment decreased to EU...	net sales of the paper segment decreased to eu...	[net, sales, paper, segment, decreased, eur, m...	[net, sale, paper, segment, decreas, eur, mn, ...	[net, sale, paper, segment, decreased, eur, mn...	2	positive
4845	negative	Sales in Finland decreased by 10.5 % in Januar...	sales in finland decreased by in january wh...	[sales, finland, decreased, january, sales, ou...	[sale, finland, decreas, januari, sale, outsid...	[sale, finland, decreased, january, sale, outs...	2	neutral

Try Pitch

OUTPUT

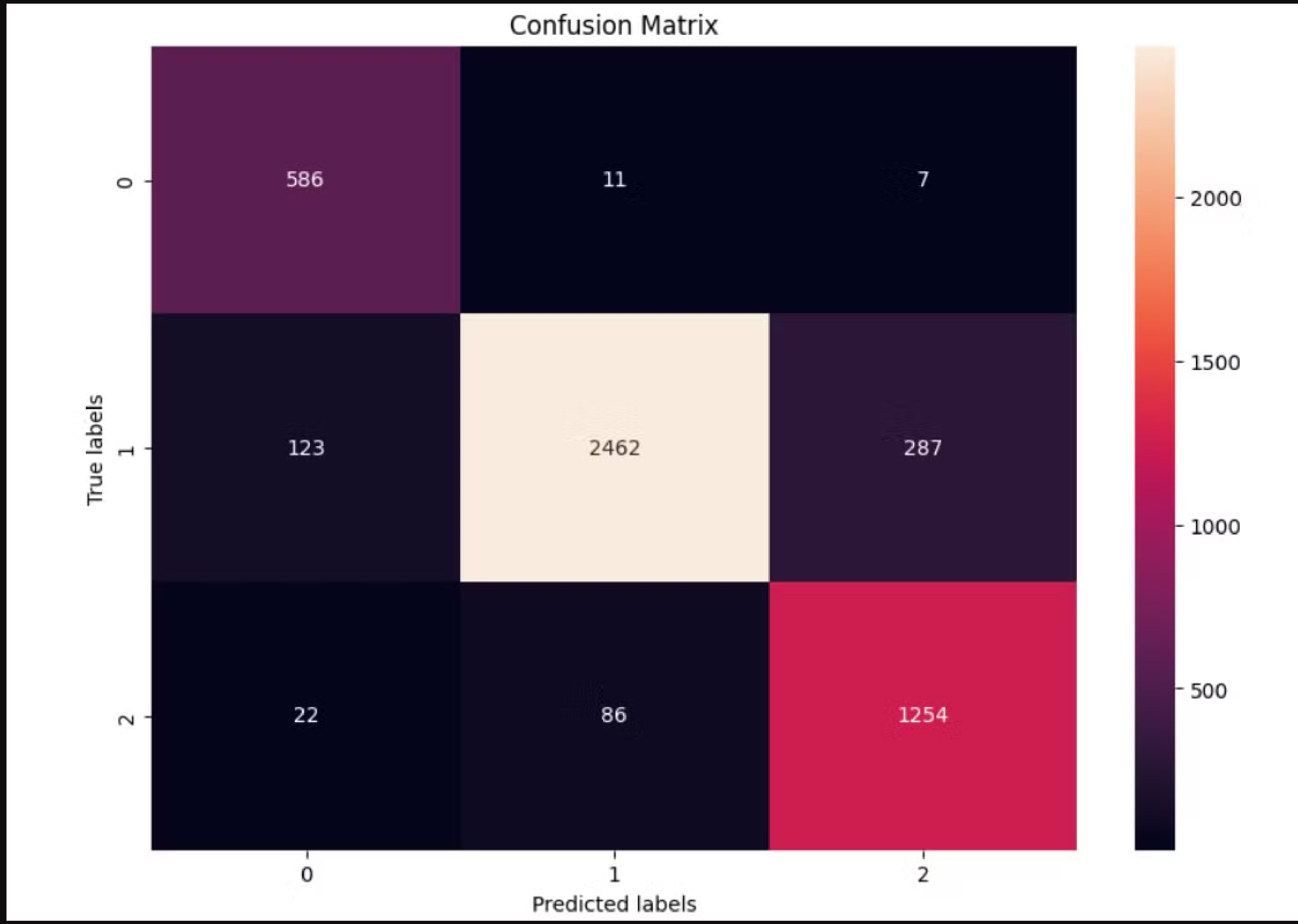
```
[I 2024-07-23 07:34:37,930] A new study created in memory with name: no-name-a165d5ae-6f10-4827-9989-2825f2b3651d
tokenizer_config.json: 100% 252/252 [00:00<00:00, 13.0kB/s]
config.json: 100% 758/758 [00:00<00:00, 46.5kB/s]
vocab.txt: 100% 232k/232k [00:00<00:00, 9.00MB/s]
special_tokens_map.json: 100% 112/112 [00:00<00:00, 7.95kB/s]
pytorch_model.bin: 100% 438M/438M [00:01<00:00, 339MB/s]
Map: 100% 3870/3870 [00:02<00:00, 1696.52 examples/s]
Map: 100% 968/968 [00:00<00:00, 2227.28 examples/s]
```

Epoch	Training Loss	Validation Loss
1	0.294600	0.289935
2	0.174100	0.363097
3	0.157200	0.387826

```
[61/61 00:29]
[I 2024-07-23 08:39:58,881] Trial 1 finished with value: 0.4986908435821533 and parameters: {'learning_rate': 4.0890084110975376e-05, 'batch_size': 16
Best trial: [0.372204065322876]
Best hyperparameters: {'learning_rate': 1.0214224747879992e-05, 'batch_size': 16, 'num_train_epochs': 3, 'weight_decay': 0.07713096613979999}
Map: 100% 3870/3870 [00:00<00:00, 10582.14 examples/s]
Map: 100% 968/968 [00:00<00:00, 7922.55 examples/s]
```

Try Pitch

OUTPUT



Try Pitch

Want to make a presentation like this one?

Start with a fully customizable template, create a beautiful deck in minutes, then easily share it with anyone.

Create a presentation (It's free)