APPLIED AI

TEXT CLASSIFICATION

Financial Sentiment Analysis

BERT and RNN Models On Financial Dataset

Dataset:

Financial Phrase Bank –v1.0

Reference Paper:

Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers

Contents

Introduction:

AI has renovated many industries, with its applications spreading from healthcare to finance. One of the major progress in AI is Natural Language Processing (NLP), which allow machines to understand and process human language. Financial sentiment analysis comes under NLP. It includes inspecting financial News to extract sentiments that can effect market trends. This project targets to calculate the efficiency of different deep learning models especially BERT and RNN, in classifying sentiments in financial news. Sentiment Analysis models proposes a way to extract actionable signals from financial news. However this job is challenging due to the domain specific language used in finance and the limited availability of large labelled datasets. General sentiment analysis models often fail to perform efficiently when applied to financial domain. To deal these challenges, this project assess the performance of two different deep learning models. Bidirectional Encoder representations from Transformers (BERT) and Recurrent Neural Network (RNN).

The main objective of research is to identify the difference in results of these two models BERT and RNN in terms of accuracy and consistency for financial analysis. The anticipated outcome is to determine the superior model and highlight the implications of using advanced AI techniques in finance.

Background:

Sentiment analysis which is also known as opinion mining, includes the use of NLP to recognize and extract particular information from the written data. In the domain of finance, sentiment analysis focus to know the sentiment expressed in financial news, reports and other written data sources. Financial Sentiment Analysis has got attention due to its quality to provide insights for investment decisions. Tradition sentiment analysis models, such as lexicon based method, have some sort of limitations in understanding the situation of financial texts. Recent progress in NLP, especially transformer models like BERT have shown promise in capturing contextual written information.

This assignment builds on existing research by contrasting BERT and RNN models. The literature on sentiment analysis is very vast and extensive and it covers many domains like finance, social media etc. Early methods to sentiment analysis basically depend on on lexicon based methods are direct and interpretable, they often lack the sophistication needed to handle the complexity and touch of financial language. With the invention of machine learning models more progressive and new methods, techniques have been developed. Machine Learning models like SVMs and logistic regression have been applied to sentiments analysis. These models uses features or functions taken out from written data e.g. frequency of words, frequency inverse document frequency (TF-IDF) use for classification of sentiments. Although machine learning models beat the lexicon based models, they still find it difficult to deal with domain-specific language of finance.

Research Gaps:

While there has been important development in sentiment analysis for finance, several research gaps remain. Firstly, there is a need for more complete assessments of different models on financial datasets to understand their strengths and weaknesses better. Most researches and studies mainly focus on single model or some limited comparisons, leaving a gap in full assessments. Secondly the effect of data processing methods and selections of constraints on model performance is often not explored. This assignment focuses to fill the following gaps by providing detailed evaluation of BERT and RNN models for sentiment analysis in finance. By comparing these models on openly available finance datasets and assessing their performance across various metrics, this study will contribute a deeper understanding of their applicability in financial domain.

Objectives:

The main and primary goal of this assignment is to compare the performance of BERT and RNN models on financial sentiment analysis. These objectives include

1. Implement the BERT and RNN model on financial dataset.
2. Training these models on financial data set.
3. Evaluate the accuracy, precision, recall and F1 score.
4. Comparing metrics of these both models.
5. Assessing the ethical implications of using AI for financial sentiment analysis. Which includes attainable, time bound and realistic, ensuring a comprehensive evaluation with the assignment time frame.
6. Provide visual such as graphs, tables to show the performance comparison of these two models.
7. Conclude the results in the context of financial sentiment analysis and discuss important and key reasons for these obtained performance differences.

Methodology:

There are some steps involves in the methodology I used. It comprises the datasets we used, how we processed it, models selections, training steps, evaluation metrics and hyperparameters tuning steps.

1. Dataset

The dataset we have used is “The Financial Phrase-bank” dataset. It is labelled by financial experts and we have used it to train and evaluate our models. This data contains the sentences from financial news articles which are then categorized into negative, positive and neutral sentiments.

2. Data Processing
First of all we have cleaned the data by removing stop words, special characters or any ambiguous numbers. Then we did tokenization, converted our written textual data into tokens using tokenizer for BERT and simple tokenizer for RNN model. Our data was in single Colum so we converted into tabular form for cleaner look and easy to understand. After that we had split the data for training, validation and for testing.

3. Model Architecture
1. BERT Model
   - BERT pre trained model with classification head & fine tune the BERT model on financial sentiment dataset.

2. RNN Model
   - The RNN architecture with the long short term memory (LSTM) layers which is followed by solid layer of classification. Trained the RNN model from scratch on the financial sentiment dataset.

4. Training procedure
Train the both models for the specific numbers of epochs, using early stopping to prevent overfitting.

5. Evaluation Metrics
Precision, Accuracy, recall, F1 score and confusion Metrix will be used to evaluate the results of both models we used.

6. Hyperparameter Tuning
We can use Grid search or random search for the hyperparameter tuning of these models for the optimal settings.

7. Visualization
By using plots and graphs we can visualize the training process, evaluation metrics and confusion metrics.

Experiments:

Here we will explain the experimental steps which includes dataset that is used, the data pre processing steps involved, configurations made for the training of models, evaluation metrics and at last the results that we achieved from our experiments. The main and primary goal of these experiments was to compare the performance of the BERT and RNN models on the financial sentiment analysis.

This experiment consists of the training of both models with specific hyperparameters. For BERT, a learning rate of 2e-5 and batch size of 16 are used, while with RNN model use a learning rate of 1e-3 and batch size of 32. These models that we used are trained for 3 epochs, and included early stopping element to overcome over fitting.

The evaluation metrics comprises of Accuracy, Precision, Recall, F1 score and also confusion metrics. The standard comparison for this experiment is a majority class classifier, which assigns the most frequent class to all instances.

The results are presented in graphs & tables highlighting the performance of each model. There are confusion metrics that are made to visualize the classification errors. This experiment focus on to provide clear cut comparison and to show the differences between these two models when they applied on financial sentiment dataset, identifying the strengths and weaknesses of each model.

Results

After performing the experiments and training of our models we came to find out that the BERT model gave far better results on financial sentiment analysis than RNN model. The accuracy we calculated for BERT is 91.75% while the RNN model is 66.7% accurate. The BERT model is not only ahead in accuracy but also outperforms in other evaluation metrics too.

a. BERT Results

BERT Model Evaluation Metrics Result

	Precision	**Recall**	**F1 score**
Neutral	0.96	0.94	0.95
Positive	0.89	0.85	0.87
Negative	0.78	0.96	0.86
Accuracy	0.92	0.92	0.92

Confusion metric For BERT Model

	Neutral Predicted	**Positive Predicted**	**Negative predicted**
Neutral Actual	286	14	5
Positive Actual	10	108	9
Negative Actual	2	0	51

b. RNN Results

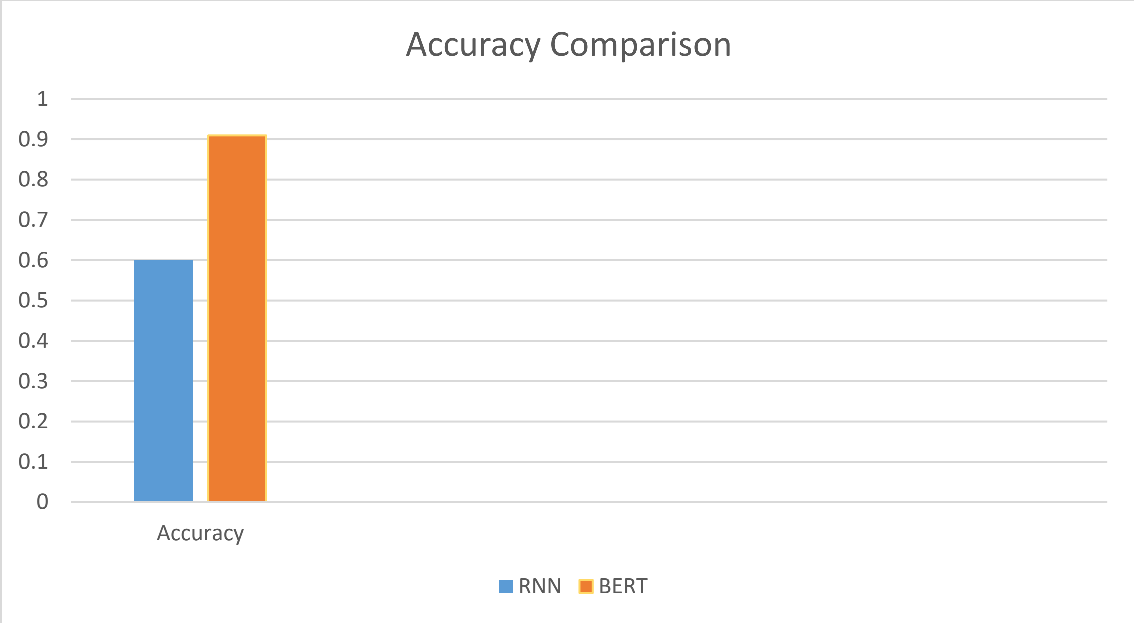
RNN Model Evaluation Metrics Result

	Precision	**Recall**	**F1 score**
Neutral	0.72	0.89	0.80
Positive	0.60	0.26	0.36
Negative	0.34	0.34	0.34
Accuracy	0.67	0.67	0.67

Confusion metric For RNN Model

	Neutral Predicted	Positive Predicted	Negative predicted
Neutral Actual	272	18	15
Positive Actual	74	33	20
Negative Actual	31	4	18

c. Comparison



These results matches with the findings of Mishev et al. 2021, highlighting the efficiency of transformer models in capturing written or contextual information. The better performance or BERT Model can be due to its ability to control the attention mechanism, which allows it to aims on relevant parts of the text. In contrast the RNN model struggle and finds it difficult to handle long term dependencies, leading to lower accuracy and higher miss classification rates.

Ethics

Ethics are the most important and critical aspect in any profession. In AI Research Ethical considerations are critical too specially when dealing with data and the algorithms that can affect the decision making process. In this financial context there are some ethical issues that must be addressed which includes data privacy, fairness and transparency.

1. Data privacy

The financial news articles and data that is used in this research or assignment are publicly accessible and do not contain any personal information about people. However, we must have to take care that any exclusive information from financial organizations is not improperly disclosed. Furthermore the methods or techniques used to collect data and process data must fulfil the relevant data protection regulations, such as General Data Protection Regulation GDPR in the European Union.

2. Bias & Fairness

In sentiment analysis the biasness can lead to unfair or misleading results, which may cause problems in financial context where decision based on these results can have important economic consequences. In this assignment I made extra efforts to reduce the biasness as much as I can by using diverse and representative dataset which is labelled by financial experts. Biasness can occur due to following reasons or can arise from several sources.

- Biasness in labelling
- Collecting biased data
- Developing biased model

3. Transparency

In AI transparency means developing a model in such a way that the decision that the model makes are understandable to users. This is specifically important in the financial background, where the users and the stake holders need to understand the sentiments to make their decisions and reach conclusions.

In this assignment I clearly documented the architectures and training methods of both BERT and RNN models. There are several techniques like SHAP (Shapely Additive Explanations) or LIME (Local Interpretable Model agnostic Explanations) can be used to explain and describe predictions of individuals. For example, emphasizing which words or phrases in a news article contributed most to the predicted sentiment.

4. Ethical Use of AI in Finance

In Finance, the applications of AI must be approached with precautions because of their significant impact on economics. Which may include.

- *Responsible Investment Decisions:* AI is used to help, support and make your work easier not to replace human decision making.
- *Regulatory Compliance*: The AI application that are used in financial domain must comply with finical standards and regulations, making sure that models we have trained do not involve or participate in un ethical practices such as manipulating the market etc.
- *Accountability*: There must be proper accountability procedures and mechanisms that must be established to deal any harm or chaotic outcomes resulting from our AI model predictions.

Conclusion

This Research based assignment represents the superior performance of BERT as compared to RNN model in financial sentiment analysis. The results that we achieved shows the effectiveness of transformer models in collecting and calculating the contextual data or written data, making them well prepared for the analysis of financial news. Along with the technical calculations we also ensure the ethical considerations in AI research, especially in financial applications. In future work we can expand the datasets, exploring the working of hybrid models that how they behave and give results, & implement explainable AI techniques to further improve model transparency and performance. The results that I found contributes to the AI field by giving comparative analysis of different deep learning models, showing their drawbacks and strengths and constraints or limitations in financial sentiment analysis.

References

Reference paper:

K. Mishev, A. Gjorgjevikj, I. Vodenska, L. T. Chitkushev and D. Trajanov, "Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers," in IEEE Access, vol. 8, pp. 131662-131682, 2020, doi: 10.1109/ACCESS.2020.3009626.

keywords: {Sentiment analysis;Feature extraction;Analytical models;Machine learning;Dictionaries;Semantics;Sentiment analysis;finance;natural language processing;text representations;deep-learning;encoders;word embedding;sentence embedding;transfer-learning;transformers;survey},

Dataset:

https://www.researchgate.net/publication/251231364_FinancialPhraseBank-v10