Financial Sentiment Analysis Using Classical and Transformer-Based Models

Nazib Ahmed, Md Azim Khan

Motivation

This project explores the effectiveness of classical machine learning models and transformer-based models in detecting sentiment from financial news headlines. The primary motivation is to assess how well traditional methods, such as TF-IDF + Logistic Regression, can interpret domain-specific financial language compared to advanced pre-trained models like BERT and FinBERT.

Financial news often contains subtle sentiment, market signals, and company performance indicators that simpler models may struggle to detect. While classical models are easier to deploy and interpret, models like FinBERT, which are fine-tuned on financial data, can capture the context and nuances of financial text better, leading to higher accuracy.

We aim to identify the model that provides the most reliable predictions for real-world applications in financial sentiment analysis while also considering the practical implications for businesses and analysts in terms of ease of use, interpretability, and accuracy.

Project Overview

1 Data Preprocessing

Clean and normalize financial headlines

- Convert text to lowercase
- Remove punctuation and extra whitespace.
- Expand financial acronyms (e.g., ROI, Q1, EBITDA)
- Replace financial units (e.g., `$10M` → `dollar10M`)

2 Class Imbalance Handling

- Subsampling the neutral class to reduce bias
- Improves model sensitivity to minority classes (positive and negative).

3 Feature Representation

Convert cleaned text into numeric vectors

- **TF-IDF**: for classical Machine Learning models.
- **Vanilla BERT**: extract `[CLS]` embeddings
- **FinBERT**: domain-specific sentiment inference

4 Model Training

Train both classical and deep learning models

- Logistic Regression, SVM, Random Forest on TF-IDF
- Fine-tuned BERT on tokenized headlines
- Direct inference using pretrained pipeline (no fine-tuning).

5 Hyperparameter Tuning

Find the optimal settings for each model

GridSearchCV is used to systematically test combinations of parameters like:

- `C`, `solver` (Logistic Regression)
- kernel, gamma (SVM)
- n_estimators, max_depth (Random Forest)

6 Model Evaluation

Evaluate model performance

- Matrics: Accuracy, Precision, Recall, F1-score(macro
- Confusion matrices highlight class-specific performance
- Analyzed where each model struggles or excels (e.g., false positives)

7 Model Comparison

Benchmark all models to select the best

- Our fine-tuned BERT model achieved 84% accuracy and 0.83 F1-score, outperforming all classical models.
- It came close to FinBERT (89.9% accuracy), despite not being pretrained on financial text.
- Classical models show decent but lower performance — especially with class imbalance

Datasets

Financial News Headlines Dataset The *Sentiment Analysis for Financial News* [1] dataset from Kaggle consists of approximately 5,000 financial news headlines, each annotated with one of three sentiment labels: **Positive, Neutral, or Negative**. The dataset provides a diverse range of short financial texts relevant to markets, earnings, and corporate events.

Custom Label Encoding:

- **Positive** → **2**
- **Neutral** → **1**
- **Negative** → **0**

In this project, we split the dataset into 70% training and 30% testing sets to evaluate both classical and transformer-based models.

Results

Table 1: Comparison of Test Accuracies Across Different Models

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.7600	0.7247	0.7170	0.7207
SVM	0.7586	0.7100	0.7249	0.7166
Random Forest	0.7428	0.7193	0.6473	0.6695
BERT (Fine-tuned)	0.8487	0.8238	0.8462	0.8343
FinBERT	0.8989	0.8744	0.9127	0.8916

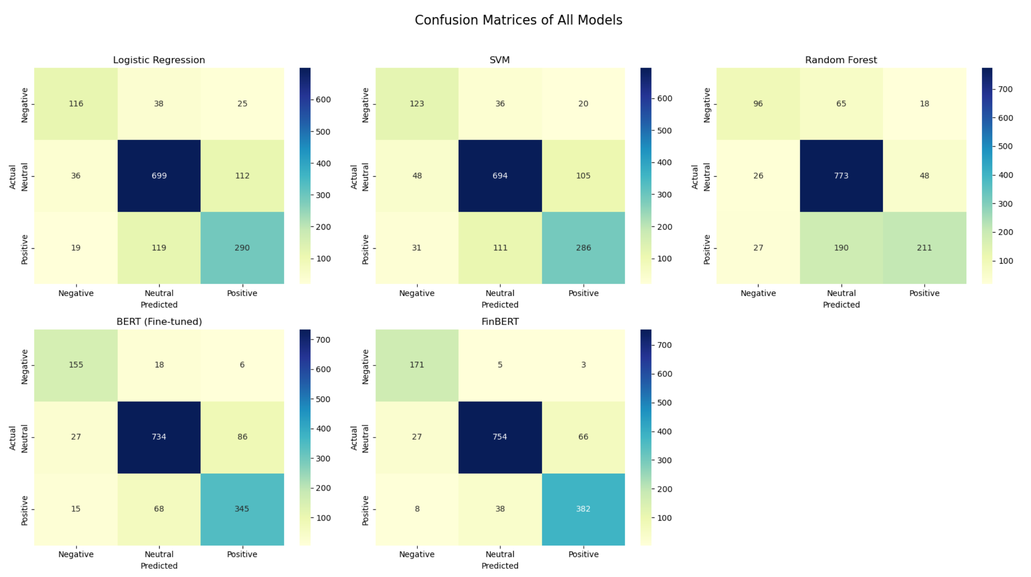

Figure 1: Confusion matrices

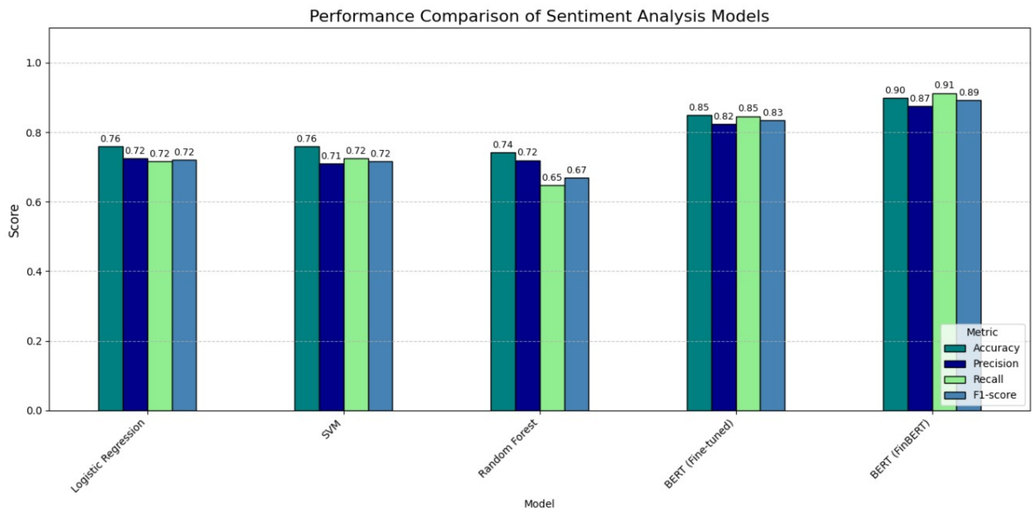

Figure 2: Performance comparison of all the models

Key Findings – Model Comparison

- Classical models (Logistic Regression, SVM, Random Forest) performed decently, especially for neutral sentiment, but struggled with positive vs negative classification.
- Fine-tuned BERT significantly outperformed all classical models
- FinBERT delivered the best overall results with 89.9% accuracy, benefiting from financial-domain pretraining.
- BERT (Fine-tuned), despite not being domain-specific, closely matched FinBERT's performance highlighting the value of fine-tuning on task-specific data.
- These results confirm the superiority of transformer-based models for financial sentiment analysis and validate BERT's reliability even without financial pretraining.

Acknowledgement

We sincerely thank our supervisors, Prof. Dr. Jörn Hees and M.Sc. Tim Metzler, for their support throughout this project. Their help was essential in deepening our understanding and knowledge of the subject.

References

[1] Ankur Z. Sentiment analysis for financial news, 2020.

[2] Pekka Malo, Ankur Sinha, Pekka Korhonen, Jyrki Wallenius, and Pyry Takala. Good debt or bad debt: Detecting semantic orientations in economic texts. *Journal of the Association for Information Science and Technology*, 65(4):782–796, 2014.

Contact

Nazib Ahmed
Hochschule Bonn-Rhein-Sieg
Email: nazib.ahmed@smail.inf.h-brs.de

Md Azim Khan
Hochschule Bonn-Rhein-Sieg
Email: md.khan@smail.inf.h-brs.de

b-it Bonn-Aachen International Center for Information Technology

Hochschule Bonn-Rhein-Sieg
University of Applied Sciences