Financial Sentiment Analysis on News and Twitter Data

Yuyao Wang

May 2023

yuyaow@bu.edu

Project Overview

Objective: Build machine learning models for sentiment classification using financial text data.

Approach:

- Pre-trained Model: DistilBERT from Hugging Face
- Data: Financial news and Twitter data with labeled sentiments (positive, neutral, negative)
- Goal: Predict sentiment in financial texts for market trend analysis and decision-making

Data Preparation & Modeling

Data Preparation:

- Clean and preprocess text data
- Convert sentiments to numerical labels (positive, neutral, negative)
- Split data into training and validation sets

Modeling:

- Fine-tune DistilBERT on financial data
- Monitor training and validation loss over epochs

Training, Evaluation, and Optimization

Metrics: Accuracy, training loss, and validation loss
Optimization:

- Early stopping, dropout, weight decay to prevent overfitting
- Hyperparameter tuning (learning rate, batch size)

Conclusion

- Fine-tuned DistilBERT achieved good performance in financial sentiment analysis
- Future work: Test additional models and further hyperparameter tuning

Word Cloud Comparison for Financial News and Twitter Financial Sentiment Datasets

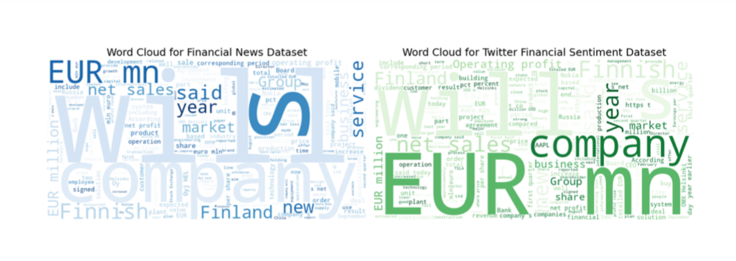


Figure 1: Word Clouds for Financial News Dataset (Left) and Twitter Financial Sentiment Dataset (Right)

Key Insights from Word Clouds

- **Financial News Dataset:**
  - Dominant terms: "EUR", "company", "net sales", and "year".
  - Focus on corporate performance metrics (net sales, profits) and international markets (mentions of countries like Finland and Russia).
- **Twitter Financial Sentiment Dataset:**
  - Dominant terms overlap with financial news but include more opinion-driven terms like "buy", "deal", and company-specific mentions (e.g., TSLA, AAPL).
  - Public sentiment reflects reactions to corporate performance and market trends.

Conclusion

- **Financial News Dataset:** Structured and fact-based, emphasizing corporate reports and performance metrics.
- **Twitter Financial Sentiment Dataset:** More opinion-driven and dynamic, highlighting public sentiment, investment behaviors, and reactions to corporate news.

Top 10 Informative Words: Financial News and Twitter Sentiment Datasets

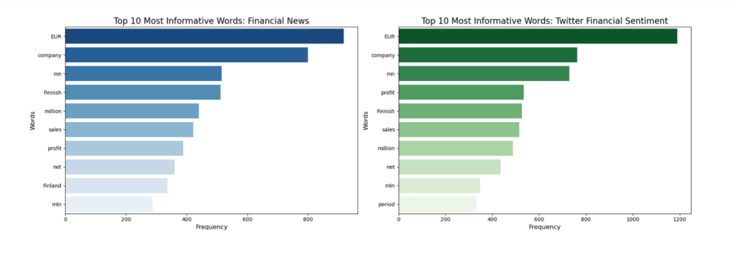


Figure 2: Top 10 Informative Words: Financial News Dataset (Left) and Twitter Financial Sentiment Dataset (Right)

Key Insights

- **Financial News Dataset:**
  - Frequent mentions of "EUR", "mn" (million), and corporate performance metrics like "sales" and "profit".
  - Focus on Finland and Eurozone, indicated by terms like "Finnish" and "Finland".
- **Twitter Financial Sentiment Dataset:**
  - Similar top words related to corporate performance and financial terms, but more conversational with words like "period".
  - Twitter discussions often revolve around specific financial periods or market reactions.

Comparison and Conclusion

- **Overlap in Topics:** Both datasets share key terms like "EUR", "company", and "profit", suggesting that Twitter users react to similar financial topics covered in news.
- **Geographic Focus:** Strong focus on Finland and the Eurozone in both datasets, likely driven by European financial events.
- **Corporate Metrics:** Heavy focus on corporate performance in both datasets, emphasizing metrics like "sales" and "profit".
- **Differences:** Twitter Financial Sentiment includes more dynamic, conversational language such as "period", indicating discussions around specific financial periods or events.

Sentiment Distribution & Proportion: Financial News vs Twitter Sentiment

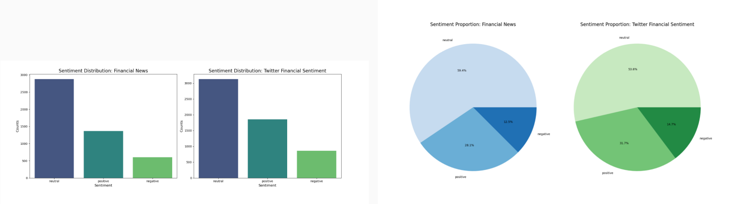


Figure 3: Sentiment Distribution Comparison

Key Insights from Sentiment Distribution & Proportion

- **Dominance of Neutral Sentiment:** Both datasets show neutral sentiment dominance, reflecting objective and fact-based content.
- **Higher Positive Sentiment on Twitter:** Twitter shows a higher positive sentiment, indicating more conversational and optimistic expressions.
- **Negative Sentiment:** While relatively low in both datasets, Twitter has a higher negative sentiment, reflecting the platform's more opinionated and reactive nature.

Conclusion

- **Similar Sentiment Trends:** Both datasets predominantly reflect neutral sentiment with positive sentiment as secondary.
- **Slight Differences:** Twitter's slightly higher negative sentiment reflects its more opinion-based content compared to the fact-driven nature of financial news reporting.

Model Construction Overview

- **Model Selection:** Pre-trained DistilBERT from Hugging Face for text classification.
- **Tokenization:** Convert text into token IDs using DistilBERT's tokenizer.
- **Data Preparation:** Preprocess and tokenize financial news and tweets, with padding/truncation for uniformity.
- **Fine-tuning:** Use Hugging Face's Trainer API to fine-tune the model on labeled sentiment data.
- **Evaluation:** Monitor model performance on validation set with early stopping to avoid overfitting.
- **Prediction:** Apply the fine-tuned model to predict sentiment in new financial texts.

Key Takeaways

- Efficient model construction using DistilBERT and Hugging Face's transformers library.
- Fine-tuning ensures model accuracy and optimal performance for financial sentiment classification.
- Early stopping prevents overfitting, enhancing model generalization to unseen data.

Training and Evaluation Results

Epoch	Training Loss	Validation Loss	Accuracy
1	0.404800	0.389999	0.839009
2	0.188400	0.451516	0.840041
3	0.170000	0.548960	0.860681

[61/61 00:04]

[879/879 04:16, Epoch 3/3]

Epoch	Training Loss	Validation Loss	Accuracy
1	0.497600	0.312796	0.841745
2	0.235700	0.406284	0.828058
3	0.224700	0.435910	0.832335

[74/74 00:05]

```
{'eval_loss': 0.43591004610061646,
 'eval_accuracy': 0.8323353293413174,
 'eval_runtime': 5.5443,
 'eval_samples_per_second': 210.849,
 'eval_steps_per_second': 13.347,
 'epoch': 3.0}
```

Figure 4: Training and Evaluation Metrics for Model 1 and Model 2

Model 1 Performance

- **Training Loss:** Consistently decreased from 0.4048 to 0.1700, indicating effective learning on the training set.
- **Validation Loss:** Dropped initially to 0.3900, but increased to 0.5489 by the third epoch, indicating potential overfitting.
- **Validation Accuracy:** Improved across epochs, rising from 83.90% to 86.07%, showing more confident predictions despite overfitting signs.

Model 2 Performance

- **Training Loss:** Decreased steadily from 0.4976 to 0.2247, reflecting efficient error minimization.
- **Validation Loss:** Initially dropped to 0.3128, but later increased to 0.4359, indicating a similar overfitting pattern as Model 1.
- **Validation Accuracy:** Dropped slightly from 84.17% to 83.23%, suggesting that generalization peaked early in the training.

Key Observations

- **Overfitting:** Both models showed signs of overfitting after the first epoch as validation loss increased while training loss continued to decrease.
- **Performance Stability:** Despite rising validation loss, both models achieved relatively high validation accuracy, with Model 1 reaching 86.07% and Model 2 at 83.23%.
- **Efficiency:** Evaluation runtime performance was efficient, processing 210 samples per second and 13 steps per second during evaluation.

Analysis of the Result Prediction Output

Input Text

"The stock market is recovering after a period of downturn."

Predicted Sentiment

2

Class Mapping

{'negative': 0, 'neutral': 1, 'positive': 2}

Sentiment Interpretation

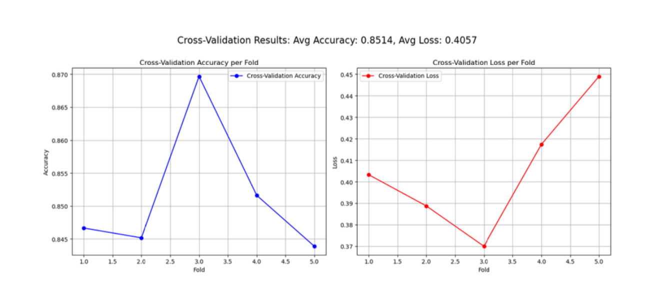
The predicted sentiment corresponds to a positive sentiment.

Next Steps to Address Overfitting

- **Regularization Techniques:**
  - **Early Stopping:** Stop training when validation loss increases.
  - **Dropout:** Introduce dropout layers to reduce overfitting.
  - **Weight Decay:** Apply L2 regularization to penalize large weights.
- **Learning Rate and Epoch Tuning:**
  - Adjust learning rate for better convergence.
  - Reduce epochs to prevent overfitting and fine-tune performance.
- **Cross-Validation:**
  - Use cross-validation for consistent model performance and reduced overfitting.

By applying these, we aim to improve the models' generalization ability and enhance overall performance on new data.

Cross-Validation Results and Insights



- **Accuracy:** The model demonstrates consistent accuracy across folds, ranging from 0.84 to 0.87. The average accuracy is **85.14%**, indicating solid performance.
- **Loss:** The average loss is **0.4057**, with variation across folds. While the loss increases in the last two folds, overall performance remains strong.

Cross-Validation Results and Insights

Potential Improvements:

- Balance the dataset splits to reduce fold variability.
- Fine-tune hyperparameters for improved stability across folds.
- Introduce advanced regularization techniques to reduce loss variance.