Stock Market News Sentiment Analysis

1/18/2026
Project 1 and UAT Gen AI Course

Contents / Agenda

Executive Summary

- Objective: Use AI to extract market sentiment from stock-related news to support investment decisions
- Approach: Compared Word2Vec and Sentence Transformer embeddings with Random Forest and Neural Network models
- Evaluation: Time-based 80/20 split with F1 score to account for class imbalance and assess generalization
- Key Result: Sentence Transformer + Neural Network achieved strongest test performance (F1 ~ 0.75) with stable generalization
- Business Value: Scalable sentiment signal to augment analyst judgment without lookahead bias

Business Problem Overview and Solution Approach

- Problem
  - Stock prices are influenced by both fundamentals and market sentiment embedded in news.
  - Manual analysis is slow and inconsistent, limiting analysts' ability to handle large news volumes and extract useful sentiment signals
- Solution Approach
  - AI-driven sentiment analysis pipeline automates sentiment extraction from large volumes of news
  - Models classify sentiment as Positive, Neutral, or Negative

EDA Results

- Key Observations from EDA
  - Sentiment labels are imbalanced, with Positive dominating and Neutral underrepresented
  - News volume varies over time, introducing potential noise that can affect sentiment signals
  - Bivariate analysis shows price movements do not react uniformly to news; sentiment impact prices probabilistically rather than deterministically.
  - Findings justify use of robust metrics beyond accuracy (F1, precision, recall)

- Business implication

  Sentiment should be treated as a supporting signal, not a deterministic trading rule

Data Preprocessing

- Checked for duplicates and inconsistencies
- Applied a time-based 80/20 split using 2019-03-25 as the cutoff
- Ensured the model trains only on past data and is evaluated on future unseen periods to prevent lookahead bias and simulate real-world deployment
- Separated predictors and target variable for modeling
- Text was transformed using two embedding strategies:
  - Word2Vec (Average of word-level embeddings)
  - Sentence Transformers (Context-aware sentence-level embeddings)
  - Sentence Transformers better capture semantic meaning in financial news

Data Preprocessing

- Built and evaluated:
  - Random Forest models
  - Neural Network models
- Used Accuracy, Precision, Recall, and F1 score as evaluation metrics due to class imbalance

Sentiment Analysis - Model Evaluation Criterion

- Evaluation Metric Selection
  - Accuracy: Measures overall correctness but can be misleading under class imbalance
  - Precision: Assesses reliability of positive sentiment predictions
  - Recall: Measures ability to capture true sentiment signals
  - F1 Score: Harmonic mean of precision and recall

- Primary Metric Chosen: F1 Score
  - Sentiment classes are imbalanced, with Neutral underrepresented
  - Accuracy alone favors the majority class and overstates performance
  - F1 balances precision and recall, capturing false positives and false negatives
  - Model selection prioritized out-of-sample F1 and stable generalization, reflecting robustness and bias–variance balance over maximizing training accuracy

Sentiment Analysis - Model Evaluation Criterion

- Supporting Metrics
  - Accuracy, Precision, Recall
  - Confusion matrices used to analyze class-level behavior
  - Enabled identification of:
    - Positive vs Negative separation
    - Weaknesses in Neutral sentiment detection
- Compared training vs test performance to assess overfitting and generalization

Sentiment Analysis - Model Building & Evaluation

- Models Evaluated
  - Evaluated Random Forest and Neural Network models across:
  - Word2Vec embeddings
  - Sentence Transformer embeddings
- Objective
  - Compare how representation choice and model type impact sentiment classification performance
  - Identify the model with the best generalization, not just training accuracy

Sentiment Analysis - Model Building & Evaluation

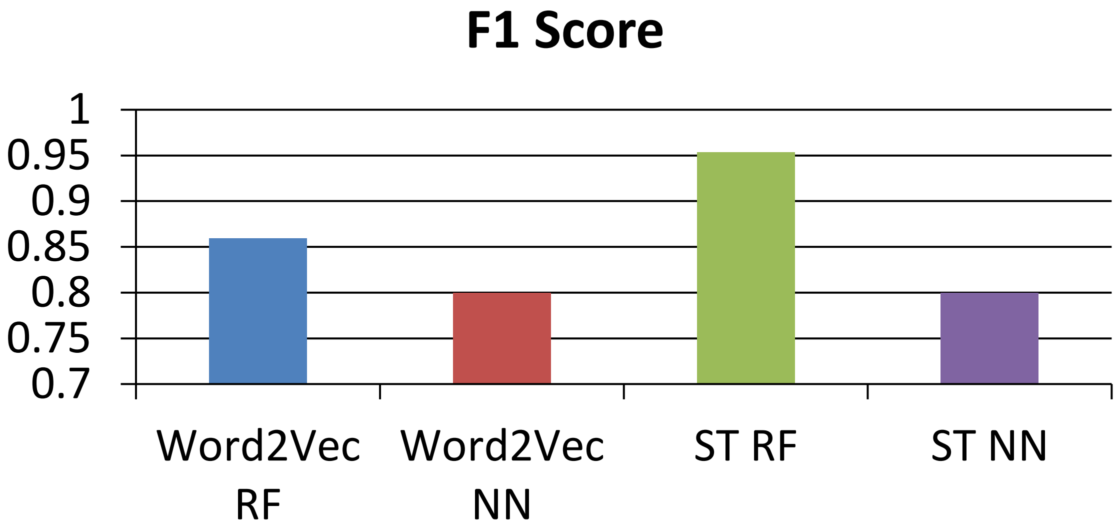
- Training Performance Summary
  - Random Forest models (Word2Vec and Sentence Transformer):
    - Exhibit very high accuracy, precision, recall, and F1 on training data (e.g., Word2Vec RF F1 ~ 0.86; Sentence Transformer RF F1 ~ 0.95)
    - Indicates strong ability to fit the training data
  - Neural Network models:
    - Exhibit noticeably lower training performance compared to Random Forests (e.g., Word2Vec NN F1 ~ 0.80; Sentence Transformer NN F1 ~ 0.80)
    - Word2Vec-based Neural Network performs weakest on training data
    - Suggest more constrained learning during training

- Key Observation
  - Random Forests exhibit overfitting; neural networks generalize better

Training Performance Comparison

	Word2Vec RF	Word2Vec NN	ST RF	ST NN
Accuracy	0.8699	0.8142	0.9628	0.8142
Recall	0.8699	0.8142	0.9628	0.8142
Precision	0.8923	0.8007	0.9459	0.8007
F1	0.8594	0.7997	0.9538	0.7997

Training F1 Score Comparison



Sentiment Analysis - Model Building & Evaluation

- Test Performance Summary
  - Random Forest models:
    - Experience a noticeable drop in F1 from training to test (e.g., Word2Vec RF F1 ~ 0.68; Sentence Transformer RF F1 ~ 0.64)
    - Performance is similar across Word2Vec and Sentence Transformer variants despite large differences in training scores
    - Indicates limited generalization despite strong training results
  - Neural Network with Word2Vec embeddings:
    - Shows modest improvement over training performance
    - Still underperforms relative to other models (F1 ~ 0.58; Precision notably low at ~ 0.50)

Sentiment Analysis - Model Building & Evaluation

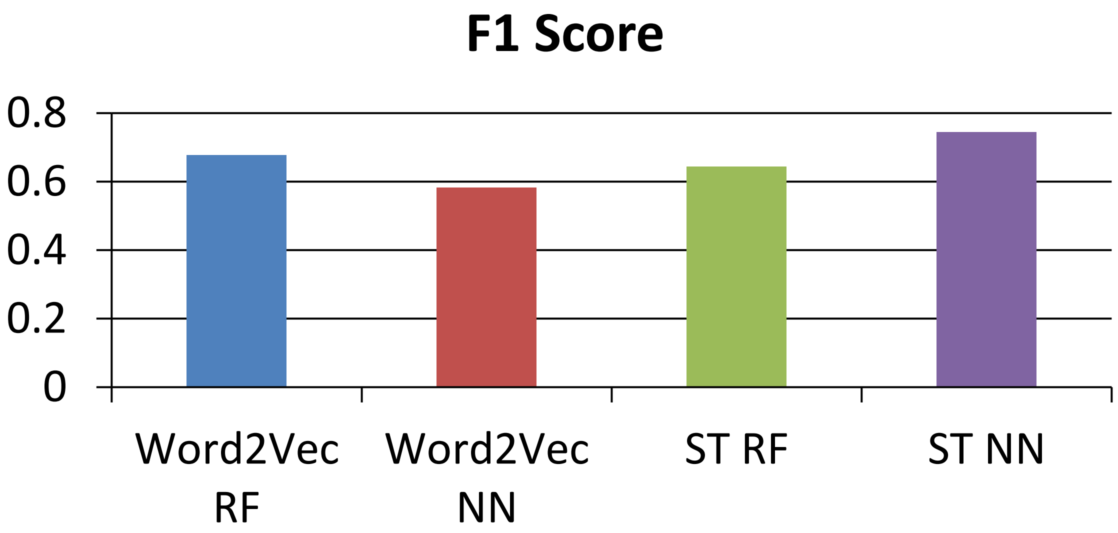
- Test Performance Summary (Cont.)
  - Neural Network with Sentence Transformer embeddings:
    - Achieves the highest test F1 and accuracy (F1 ~ 0.75; Accuracy ~ 0.77; Recall ~ 0.77; Precision ~ 0.74)
    - Outperforms all other models on unseen data
    - Test metrics exceed or closely match training metrics, indicating strong generalization

- Key Observation:
  - Sentence Transformer embeddings materially improve test-time performance
  - Neural Network benefits most from contextual sentence-level representations

Test Performance Comparison

	Word2Vec RF	Word2Vec NN	ST RF	ST NN
Accuracy	0.7158	0.7053	0.7263	0.7684
Recall	0.7158	0.7053	0.7263	0.7684
Precision	0.6777	0.4974	0.7247	0.7445
F1	0.6769	0.5834	0.6442	0.7452

Test F1 Score Comparison



Sentiment Analysis - Model Building & Evaluation

- Confusion Matrix Insights (Training vs Testing)
  - Training Confusion Matrices
    - Random Forest models display near-perfect class separation
    - Suggest memorization of training patterns rather than robust learning
  - Test Confusion Matrices
    - Sentence Transformer + Neural Network:
      - Best separation between Positive and Negative sentiment classes
      - More balanced prediction distribution across classes
  - Neutral sentiment remains hardest to classify across all models due to class imbalance and inherent ambiguity

Sentiment Analysis - Model Building & Evaluation

- Key Observations Across Models
  - Embedding choice has a larger impact on performance than model choice
  - Sentence Transformers consistently outperform Word2Vec across both model families
  - Random Forest models show high variance, with strong training performance but weaker test results
  - Neural Networks demonstrate better bias–variance balance when paired with contextual embeddings

Sentiment Analysis - Model Building & Evaluation

- Selected Model

Neural Network with Sentence Transformer embeddings

- Justification
  - Highest **out-of-sample F1 score (~ 0.75)** among all models
  - Balanced between precision and recall under class imbalance
  - Demonstrated stable generalization with a moderate train–test gap
  - Lower overfitting risk compared to Random Forest alternatives
  - Effectively leveraged contextual sentence-level embeddings
  - Well-suited for real-world deployment where robustness matters

Sentiment Analysis - Model Building & Evaluation

- Final Takeaways for the Business
  - Context-aware sentence embeddings significantly enhance sentiment analysis quality
  - Model evaluation should prioritize generalization over training accuracy
  - The selected model provides a robust and reliable sentiment signal suitable for integration into investment decision workflows
  - Sentiment should be used as a decision-support signal, not a deterministic trading rule

Insights and Next Steps

- Key Insights
  - Context-aware embeddings significantly improve sentiment modeling
  - Model complexity should be controlled to avoid overfitting
  - Neutral sentiment remains challenging due to data imbalance
- Recommended Next Steps
  - Apply additional regularization or tuning to Random Forest models to mitigate overfitting
  - Introduce validation splits and early stopping for Neural Networks
  - Explore fine-tuning pre-trained Sentence Transformer models for higher performance ceilings
  - Aggregate sentiment signals weekly to reduce noise
  - Improve Neutral sentiment detection via class weighting

Appendix: Data Background

- Dataset contains historical news articles linked to a NASDAQ-listed company
- Each record includes date, news text, and sentiment label
- Used solely for analytical and educational purposes

APPENDIX

Business Value of Implementing This Solution

- Faster interpretation of market-moving news
- More consistent sentiment signals across analysts
- Improved risk awareness during volatile periods
- Scalable framework adaptable to multiple tickers or markets

Bottom line
This solution provides measurable, explainable sentiment intelligence that enhances decision-making without overpromising predictive certainty.

Great Learning

Happy Learning !