

Abstract geometric lines in black on a white background, forming various overlapping polygons and triangles.

SENTIMENT ANALYSIS ON FINANCIAL NEWS

Submitted by
Swati Mishra

TABLE OF CONTENTS

- Overview
- Problem Statement
- Datasets
- Methodology
- Insights Report
- Expected Outcomes
- Conclusion



OVERVIEW

Sentiment analysis, also known as opinion mining, is a natural language processing (NLP) technique used to determine the emotional tone of a piece of text. Analyzing the sentiment of news articles, press releases, and market reports in the financial domain can provide valuable insights into market trends, investor confidence, and potential stock price movements. By leveraging machine learning and deep learning methods, sentiment analysis on financial news can assist investors, analysts, and policymakers make informed decisions.

For this assignment, I will combine multiple approaches taught in class. This hybrid approach will help me gain deeper insights into financial narratives shaping market outcomes.

- **Lexicon-based method (VADER)** for quick and interpretable sentiment scoring.
- **Topic Modeling (LDA)** to uncover hidden themes in financial news.
- **Machine Learning and Deep Learning models** to improve classification accuracy and capture financial domain-specific context.



PROBLEM STATEMENT

Financial markets are highly sensitive to information flow, but the volume and complexity of financial news make manual analysis unfeasible.

Traditional sentiment analysis tools often fail to capture financial jargon and underlying themes.

The challenge is to design an **automated, robust, and accurate pipeline** that combines lexicon-based sentiment detection (positive, negative, neutral), topic modeling, and machine/deep learning techniques while ensuring exception handling and interpretability through visualizations and insights reporting.



DATASETS DOWNLOADED

- *Kaggle Financial Phrase Bank*
<https://www.kaggle.com/dattran2346/finacial-parse-bank?resource=download>
- *Reuters Financial News Dataset*
<https://github.com/duynht/financial-news-dataset>



METHODOLOGY

Data Preprocessing

Text cleaning (stopwords, special characters, numbers)

Tokenization and lemmatization

Domain-specific dictionary handling (financial terms)

Sentiment Analysis (VADER)

Apply VADER to classify text as positive, negative, or neutral

Generate compound sentiment scores for nuanced insights

Topic Modeling (LDA - Latent Dirichlet Allocation)

Extract latent themes (e.g., inflation, stock movements, mergers)

Analyze how topics correlate with sentiment classes

Machine Learning

Feature extraction: Bag-of-Words (BoW), TF-IDF

Models: Logistic Regression, Naïve Bayes, Support Vector Machines (SVM)

Evaluate and compare using accuracy, precision, recall, and F1-score



METHODOLOGY (CONTINUED)

Deep Learning

Embeddings: Word2Vec, GloVe

Models: LSTM, Transformer-based architectures (BERT/FinBERT)

Fine-tune on labeled datasets for financial sentiment classification

Visualization

Sentiment distribution (bar, pie)

Word clouds for positive vs. negative sentiment.

Topic distribution from LDA



INSIGHTS REPORT

- Correlate sentiment with:
 - Market trends
 - Investor confidence
 - Potential stock price movements
 - Any other financial information
- Compare each approach: lexicon-based vs machine learning vs deep learning performance
- Discuss which approach works best for financial contexts



EXPECTED OUTCOMES

- A hybrid pipeline capable of:
 - Detecting sentiment in financial news
 - Identifying hidden themes via LDA
 - Improving accuracy with ML/DL models
 - Visualizations that make the results interpretable
- Insights linking news sentiment with market movements
- A robust system with **exception handling** for real-world deployment



CONCLUSION

This study demonstrates that combining multiple approaches, VADER for lexicon-based sentiment detection, LDA for uncovering latent themes, and ML/DL methods for context-aware classification, provides a comprehensive framework for analyzing financial news. It shows how sentiment analysis, when supported with topic modeling and advanced ML/DL models, can generate actionable insights into market trends, investor confidence, and potential stock price movements.

Overall, this project will deepen my understanding of applying AI techniques to finance and show me the practical challenges and opportunities in sentiment-driven market analysis.

Through this assignment, I will grasp a solid foundation in:

- How to preprocess and clean financial text data for NLP tasks.
- The strengths and limitations of lexicon-based sentiment analysis (VADER) compared to ML and DL approaches.
- How topic modeling (LDA) complements sentiment analysis by revealing hidden themes.
- The importance of visualization in making results interpretable and decision-oriented.
- How exception handling is critical for building robust, real-world NLP pipelines.



THANK YOU

Swati Mishra

+91-8726552080

mswati547@gmail.com

GitHub link: <http://github.com/mswati-tech/>

LinkedIn link: <https://www.linkedin.com/in/dr-swati-mishra-042372125/>