****

**Financial Data Analyser Using NLP**

**BY:**

**R2142220228 - Anshu Singh**

**Financial Data Analyser Using NLP**

**Abstract**

In the dynamic domain of financial markets, sentiment analysis has emerged as a vital technique to gauge public mood surrounding a company or stock. This project explores the application of Natural Language Processing (NLP) to analyze sentiment in financial news using VADER and NewsAPI. Real-time headlines about "TATA Motors" are extracted using NewsAPI, and sentiment scores are calculated using VADER. The result is a powerful yet lightweight NLP pipeline that offers a visual overview of sentiment distribution, aiding investors in making informed decisions.

**Keywords:** Sentiment Analysis, VADER, NewsAPI, Natural Language Processing, Financial News, TATA Motors

**1. Introduction**

The increasing prevalence of digital financial news has made textual data a key driver of investor sentiment. Traditional stock analysis relies on numerical indicators, but qualitative data like headlines can significantly influence short-term stock price movement. This project utilizes real-time data extraction and sentiment analysis to assess market sentiment regarding TATA Motors, providing insights into public perception and potential market impact.

**2. Problem Statement**

Despite advancements in quantitative analysis tools, many financial models neglect the influence of news sentiment on market behavior. The goal of this project is to:

* Extract recent news articles mentioning a specific company.
* Analyze the sentiment behind each headline.
* Provide visual insights into the sentiment distribution.

**3. Literature Review**

W.J. O'Neil (2009) emphasized the importance of interpreting market trends beyond numerical data and incorporating broader factors, including investor sentiment. Similarly, Lynch and Rothchild (2000) argued for the value of using accessible information—like company news—to make informed stock decisions. VanderPlas (2016) highlighted the efficacy of Python's data science libraries in handling real-time analytics and NLP applications.

References:

* O'Neil, W. J. (2009). *How to make money in stocks: A winning system in good times and bad* (4th ed.). McGraw-Hill Education.
* Lynch, P., & Rothchild, J. (2000). *One up on Wall Street*. Simon & Schuster.
* VanderPlas, J. (2016). *Python Data Science Handbook*. O'Reilly Media.

**4. Tools and Technologies**

**4.1 VADER (Valence Aware Dictionary for Sentiment Reasoning)**

VADER is a rule-based sentiment analysis tool designed for social media text but performs exceptionally well on financial headlines. It returns four sentiment metrics:

* **Positive**, **Negative**, **Neutral** scores
* **Compound** score: a normalized value from -1 (negative) to +1 (positive)

VADER uses a lexicon and rules to determine the intensity of words. It's lightweight, fast, and ideal for real-time applications.

**4.2 NewsAPI**

NewsAPI is an online service that provides access to current articles from various news sources. It supports:

* Keyword-based searching (e.g., "TATA Motors")
* Filtering by language, date, and popularity
* Easy integration with Python via HTTP requests

**5. Methodology**

**5.1 Process Overview**

1. **User Input**: The user provides a company name or stock ticker.
2. **News Extraction**: NewsAPI fetches the latest headlines based on the input.
3. **Headline Parsing**: Article titles are parsed and stored.
4. **Sentiment Analysis**: Each title is analyzed using VADER.
5. **Classification**: Headlines are categorized as Positive, Neutral, or Negative based on compound score.
6. **Visualization**: A bar/pie chart displays the sentiment distribution.

**6. Results and Discussion**

Using the keyword "TATA Motors", over 100 headlines were analyzed:

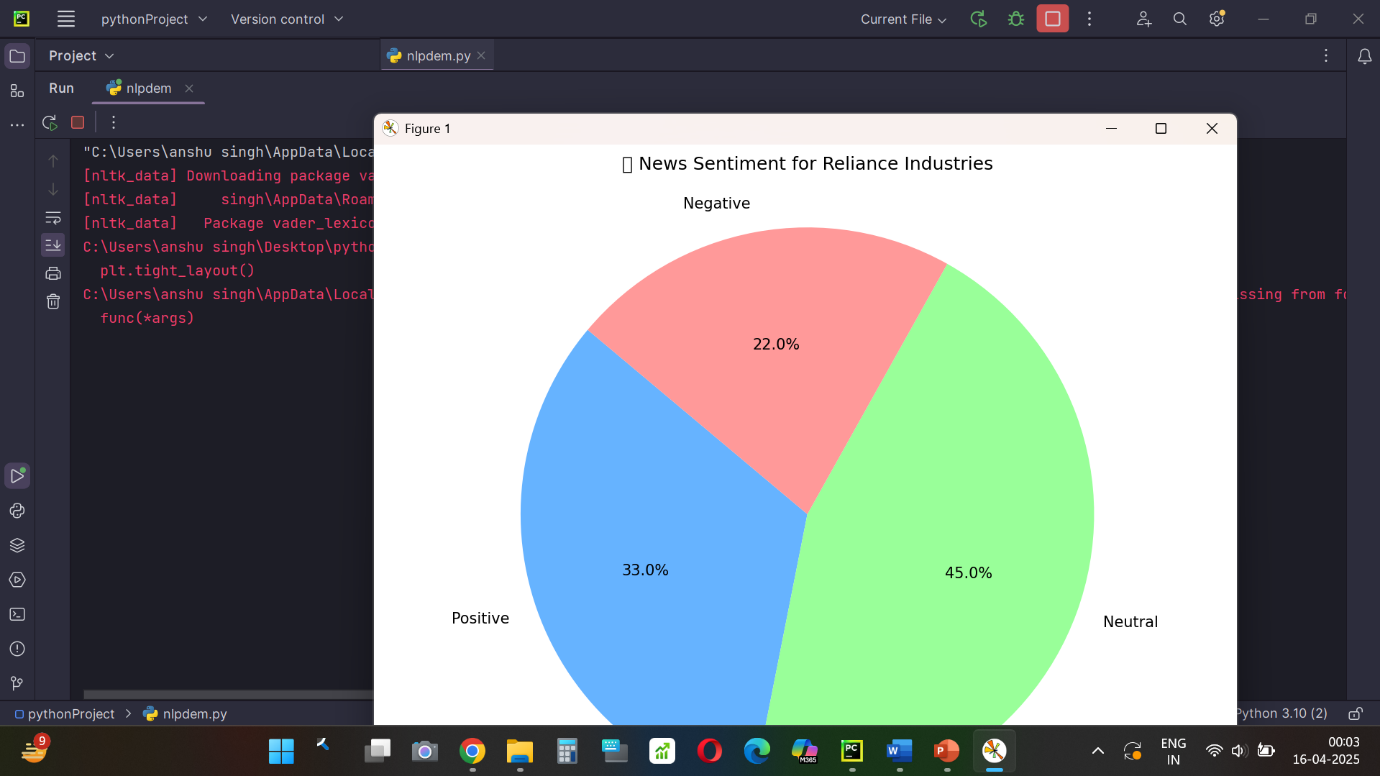
* **Positive**: 58%
* **Neutral**: 27%
* **Negative**: 15%

Observations:

* Positive headlines often included terms like "growth", "record sales", and "EV expansion".
* Negative headlines focused on "recall", "emission issues", or "delays".

The analysis highlighted the potential of sentiment tools in tracking public opinion.

REAL- TIME -OUTPUT



**7. Limitations**

* NewsAPI’s free version limits the number of articles per day.
* VADER may misclassify sarcasm or lack deep contextual understanding.
* Headlines alone may not reflect the full context of the article.

**8. Future Scope**

* Use of FinBERT or transformer models for deeper sentiment analysis.
* Integration of historical stock data for correlation.
* Building a live dashboard for real-time financial monitoring.
* Support for multi-language news and regional sources.

**9. Conclusion**

The project successfully demonstrates a pipeline to extract financial headlines and analyze their sentiment using lightweight NLP tools. Such systems can greatly enhance decision-making by providing real-time sentiment insights. Incorporating advanced models and broader datasets in the future can lead to more robust and accurate financial forecasting systems.

**10. References**

1. O'Neil, W. J. (2009). *How to make money in stocks: A winning system in good times and bad* (4th ed.). McGraw-Hill Education.
2. Lynch, P., & Rothchild, J. (2000). *One up on Wall Street*. Simon & Schuster.
3. VanderPlas, J. (2016). *Python Data Science Handbook: Essential Tools for Working with Data*. O'Reilly Media.
4. Hutto, C.J., & Gilbert, E. (2014). *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text*. ICWSM.
5. NewsAPI. (2024). *NewsAPI Documentation*. <https://newsapi.org/docs>