The Pulse of the Market

Social Media Sentiment & Stock Prices
Team: The Analytix

Project Overview

- Analyze how social media sentiment affects stock prices.
- Data sources: Twitter, Reddit, StockTwits, blogs.
- Use sentiment analysis and predictive models to identify trends.

Data Collection

- Platforms: Twitter, Reddit, StockTwits, financial blogs.
- Data format: Timestamp, post, user, stock mentioned.
- Collected using APIs and web scraping tools.

Data Cleaning & Missing Values

- Removed duplicates and irrelevant entries.
- Filtered out bot and spam accounts.
- Handled missing values using imputation and row removal.

Sentiment Analysis Techniques

- Used VADER from NLTK for scoring sentiment.
- Scores range from -1 (negative) to 1 (positive).
- Targeted over 75% classification accuracy.

Feature Selection & Engineering

- Key features: Sentiment score, post volume, topic frequency.
- Engineered features: Lag variables, moving averages.
- Correlation and variance analysis for selection.

Handling Outliers & Data Transformation

- Outliers identified using IQR and Z-score methods.
- Applied log transformations to skewed data.
- Normalized features for neural network input.

Summary Statistics

- Included mean, median, standard deviation, min, max.
- Analyzed daily averages and volatility.
- Used statistical summaries to guide modeling.

Predictive Modeling

- Models: Regression, ARIMA, LSTM.
- Metrics: R-squared, RMSE, Sharpe ratio.
- Example: LSTM model with 0.8 R-squared for AAPL.

Case Study: GameStop (GME)

- January 2021 short squeeze driven by Reddit.
- Spike in positive sentiment ahead of stock surge.
- Stock rose 1,700%, subreddit grew massively.

Case Study: Elon Musk & Tesla

- Musk's tweets often influence TSLA price.
- Example: $420 tweet led to significant movement.
- Average ±5% fluctuation after tweets.

Risks & Limitations

- Risk of market manipulation through fake sentiment.
- Data bias due to bots and fake accounts.
- SEC monitoring misleading financial promotions.

Future Trends

- AI for real-time sentiment tracking.
- Fake news detection improvements.
- Wider adoption of alternative data by hedge funds.

Conclusion

- Sentiment is a useful but imperfect investment signal.
- Blend sentiment with traditional indicators.
- Remain cautious of manipulation and bias.

GitHub & Contact

- GitHub: https://github.com/yourteam/sentiment-stock-project
- Contact: Rounakdhall0029@gmail.com
- Tools used: Python, NLTK, Pandas, Matplotlib, etc.