

Financial Market Analysis

Sankalp Setia

Indian Institute of Technology Guwahati

25 November 2024

Outline

- Introduction
- Aim of the Study
- Experimental Procedure
- Results and Analysis
- Discussion and Conclusion
- Future Work
- Bibliography

Introduction

- ▶ Being able to predict future outcomes has been a cornerstone of human curiosity and innovation.
- ▶ Among the most challenging domains of prediction lies financial forecasting— a field characterized by its unpredictability and the high stakes of its outcomes.
- ▶ This research seeks to answer the critical question: How much can sentiment truly influence prediction?

Introduction

- ▶ Through the use of regression-based models and ensemble techniques, we aim to unravel these mysteries and contribute to a deeper understanding of the factors that shape stock market dynamics.
- ▶ We explore how emotions and opinions—captured from social media platforms like Twitter—might offer additional layers of insight into market behavior.

Aim of the Study

- ▶ Evaluate the impact of sentiment analysis on financial forecasting.
- ▶ Comprehensive model evaluation: Undertake a thorough comparison of various advanced forecasting models in financial time series, and machine learning techniques.

Experimental Procedure: Overview

- ▶ Incorporates sentiment scores and traditional stock indicators.
- ▶ Analyzes 5 major stocks: TSLA, AAPL, AMZN, GOOG, NFLX.
- ▶ Creation of diverse feature sets: Unique combinations of features were developed, half including only closing prices, and half combining opening and closing prices, to thoroughly investigate the data's predictive capacity.

Close/open price	Time Series Model	Time Step
Close {Close,Open}	Linear Ridge Elastic Net Ensemble	7 Day 14 Day

Algorithms Employed

- ▶ Linear Regression: Simple and interpretable.
- ▶ Ridge Regression: Adds L2 regularization to prevent overfitting on training data.
- ▶ Elastic Net Regression: Combines L1 and L2 regularization to create a hybrid model.
- ▶ Contextual effectiveness: These algorithms were specifically selected for their proven capability in similar financial contexts and their ability to handle intricate data structures.

Algorithms Employed

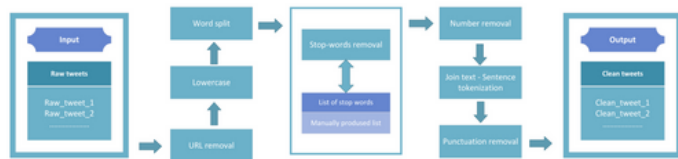
- ▶ Diverse temporal analysis: Experiments were conducted over two distinct time steps to capture both short-term and long-term market trends, reflecting the stock data's predictive patterns.
- ▶ Comprehensive model assessment: This approach enabled a thorough evaluation of the models' performance across varying temporal contexts, ensuring a more complete understanding of their predictive capabilities.

Sentiment Analysis

- ▶ Crucial role of sentiment analysis: Utilized sentiment analysis, applying this tool to Twitter data to derive sentiment scores.
- ▶ Depth with rolling averages: Incorporated 7-day and 14-day rolling averages to capture short-term and medium-term sentiment trends, providing a richer analysis of market sentiment.
- ▶ Multi-dimensional market view: Generates sentiment scores ranging from negative to positive.

Data Preprocessing

- ▶ Raw text data from Twitter often contains noise, such as URLs, numbers, and punctuation, which can reduce the effectiveness of machine learning models.
- ▶ Detailed preprocessing steps: Involved removing hyperlinks from tweets, converting text to lowercase, and breaking it into words. Extraneous phrases and numerical strings were removed, and the text was restructured into its original form with tokenization.



Metrics: Root Mean Squared Error (RMSE)

- ▶ Quantifies prediction accuracy.
- ▶ Higher weight to larger errors.
- ▶ A critical measure for comparing model performance.

$$\text{RMSE} = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}} \quad (1)$$

Experiment Execution

- ▶ Models trained on feature combinations (technical + sentiment).
- ▶ Predictions combined using ensemble weights.
- ▶ Evaluated for both short-term and long-term forecasting.

Key Findings

- ▶ Sentiment analysis gives accurate predictions.
- ▶ Ridge Regression provides stability and reduces overfitting.
- ▶ Ensemble approach consistently outperforms individual models.
- ▶ Elastic Net gives better predictions than Ridge.

Discussion: Insights

- ▶ Ridge Regression: Effective for high number of correlation features.
- ▶ Elastic Net: Balances Ridge with Lasso when not every feature is useful.
- ▶ Sentiment scores enhance model responsiveness to market events.
- ▶ Ensemble models leverage individual strengths for improved results.

Conclusion

- ▶ Sentiment analysis complements traditional indicators effectively.
- ▶ Ridge and Elastic Net enhance prediction accuracy.
- ▶ Study provides a pathway for integrating sentiment and technical data in financial forecasting.
- ▶ Ensemble Learning Efficacy: By combining Linear Regression, Ridge Regression, and Elastic Net through weighted averages, the ensemble model showcased superior accuracy and generalizability.

Future Work

- ▶ Explore additional data sources from different social media sites for sentiment analysis.
- ▶ Experiment and learn about advanced models for sequential predictions.

Bibliography

1. H. Zou and T. Hastie, “Regularization and variable selection via the elastic net,” Journal of the Royal Statistical Society: Series B, 2005.
2. L. Breiman, “Random forests,” Machine Learning, 2001.
3. VADER Sentiment Analysis Documentation.
4. R. Tibshirani, “Regression shrinkage and selection via the Lasso,” Journal of the Royal Statistical Society, 1996.

Thank You!