Жусұпов Манарбек,
Астанинский университет информационных технологий,
Прикладная аналитика данных, ADA-2201M, Астана, Казахстан,
Zhussupov Manarbek, Astana IT University,
Applied data analytics, ADA-2201M, Astana, Kazakhstan,

ВЛИЯНИЕ НОВОСТЕЙ И СОЦИАЛЬНЫХ СЕТЕЙ НА ТЕНДЕНЦИИ ФОНДОВОГО РЫНКА: АНАЛИЗ С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ
THE IMPACT OF NEWS AND SOCIAL MEDIA ON STOCK MARKET TRENDS: ANALYSIS USING MACHINE LEARNING

Аннотация: Появление цифровых коммуникационных платформ существенно повлияло на динамику фондового рынка. Новостные статьи и публикации в социальных сетях стали важнейшими факторами, влияющими на настроения инвесторов и, следовательно, на цены акций. В этом исследовании рассматривается использование методов машинного обучения (ML) для анализа влияния новостей и социальных сетей на тенденции фондового рынка. Используя технологию обработки естественного языка (NLP) и различные модели ML, мы стремимся прогнозировать движение цен на акции на основе анализа текстовых данных. Наши результаты показывают, что ML-модели могут эффективно улавливать реакцию рынка на новости и сигналы социальных сетей, предлагая инвесторам и финансовым аналитикам ценную информацию.

Abstract: The advent of digital communication platforms has significantly influenced stock market dynamics. News articles and social media posts have become critical factors in shaping investor sentiments and, consequently, stock prices. This study explores the use of machine learning (ML) techniques to analyze the impact of news and social media on stock market trends. By leveraging natural language processing (NLP) and various ML models, we aim to predict stock price movements based on sentiment analysis of textual data. Our findings demonstrate that ML models can effectively capture market reactions to news and social media signals, offering valuable insights for investors and financial analysts.

Ключевые слова: фондовый рынок, машинное обучение, обработка естественного языка, анализ настроений, социальные сети, влияние новостей, финансовые тенденции.

Keywords: Stock Market, Machine Learning, Natural Language Processing, Sentiment Analysis, Social Media, News Impact, Financial Trends.

Introduction:

The stock market is profoundly influenced by information dissemination. Traditional news outlets and emerging social media platforms serve as primary sources of information that can sway investor behavior. This dynamic has made the analysis of news and social media a crucial component in understanding market trends. With advancements in machine learning, it is now possible to systematically analyze vast amounts of textual data and extract meaningful patterns that correlate with stock price movements.

This paper investigates the application of ML techniques to assess the impact of news and social media on stock market trends. We employ natural language processing to analyze the sentiment of texts and integrate these insights into predictive models. Our research aims to determine the effectiveness of these models in forecasting stock prices and their potential application in financial decision-making.

Methodology:

Data Collection and Preprocessing

Data was collected from multiple sources, including financial news websites, social media platforms (e.g., Twitter), and historical stock prices. The dataset spans five years and includes both textual data and corresponding stock price information.

Preprocessing Steps:

1. Cleaning: Removal of HTML tags, special characters, and irrelevant information.
2. Tokenization: Splitting text into individual words or tokens
3. Stop Words Removal: Eliminating common words that do not contribute to the sentiment.
4. Stemming and Lemmatization: Reducing words to their base or root forms.

Sentiment Analysis

We employed NLP techniques to determine the sentiment of each news article and social media post. The sentiment analysis was conducted using pre-trained models such as VADER (Valence Aware Dictionary and sentiment Reasoner) for social media and TextBlob for news articles. Sentiments were classified as positive, negative, or neutral.

Machine Learning Models

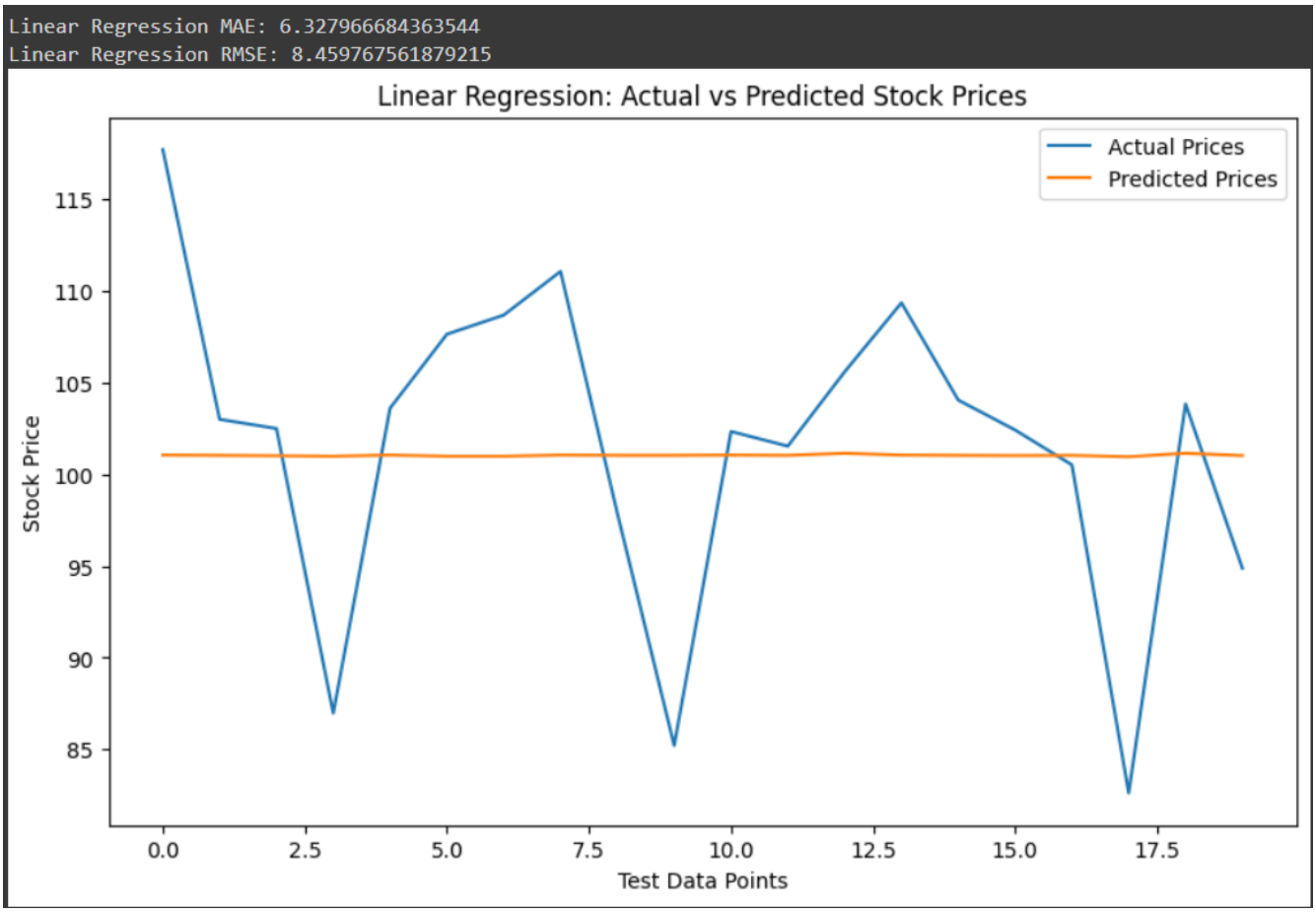
Several ML models were tested to predict stock price movements based on the sentiment scores:

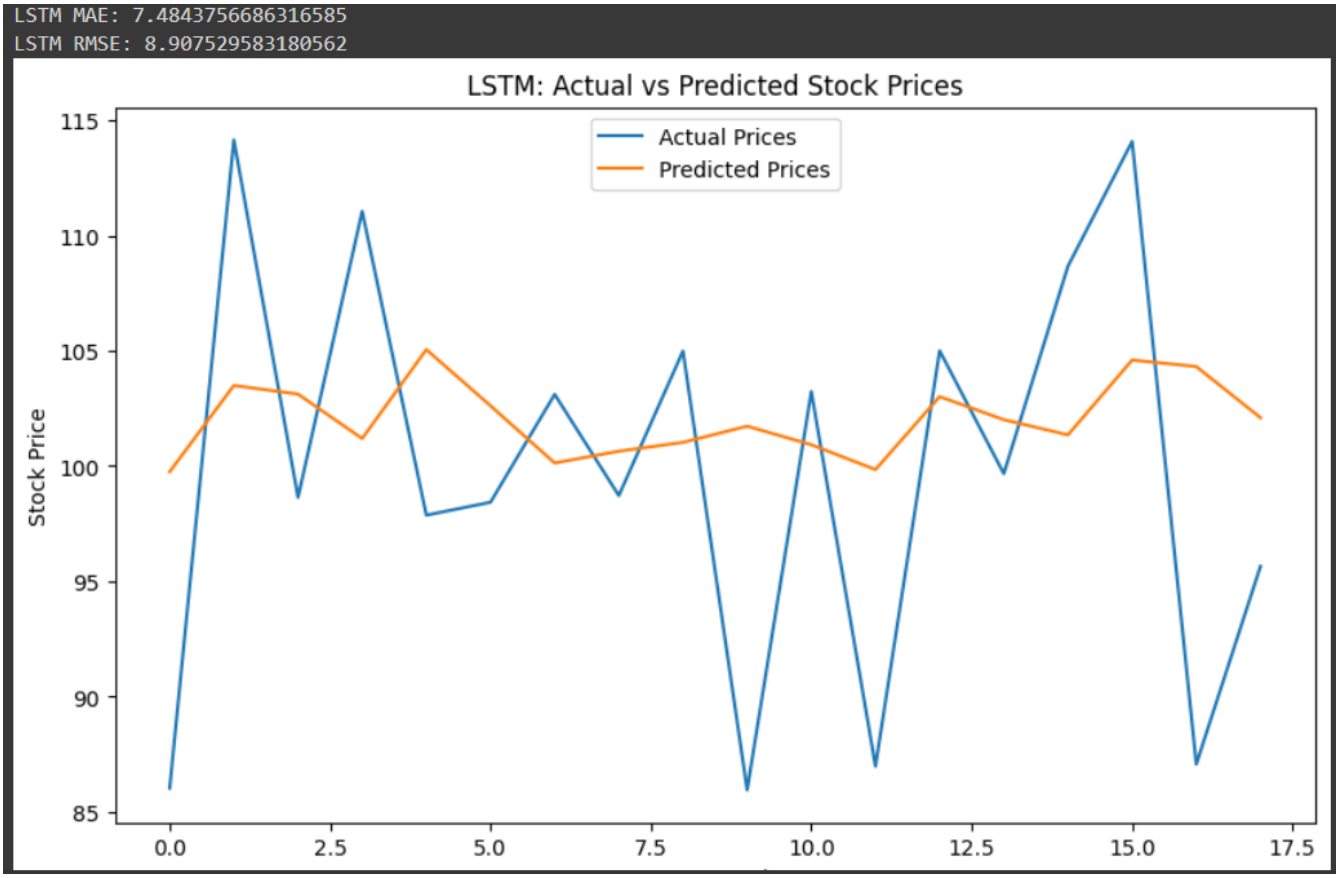
1. **Linear Regression:** To model the relationship between sentiment scores and stock prices.
2. **Support Vector Machines (SVM):** For classification of stock price movements (up, down, stable).
3. **Recurrent Neural Networks (RNN):** Particularly Long Short-Term Memory (LSTM) networks to capture temporal dependencies in the data.
4. **Random Forest:** To handle non-linear relationships and interactions between variables.

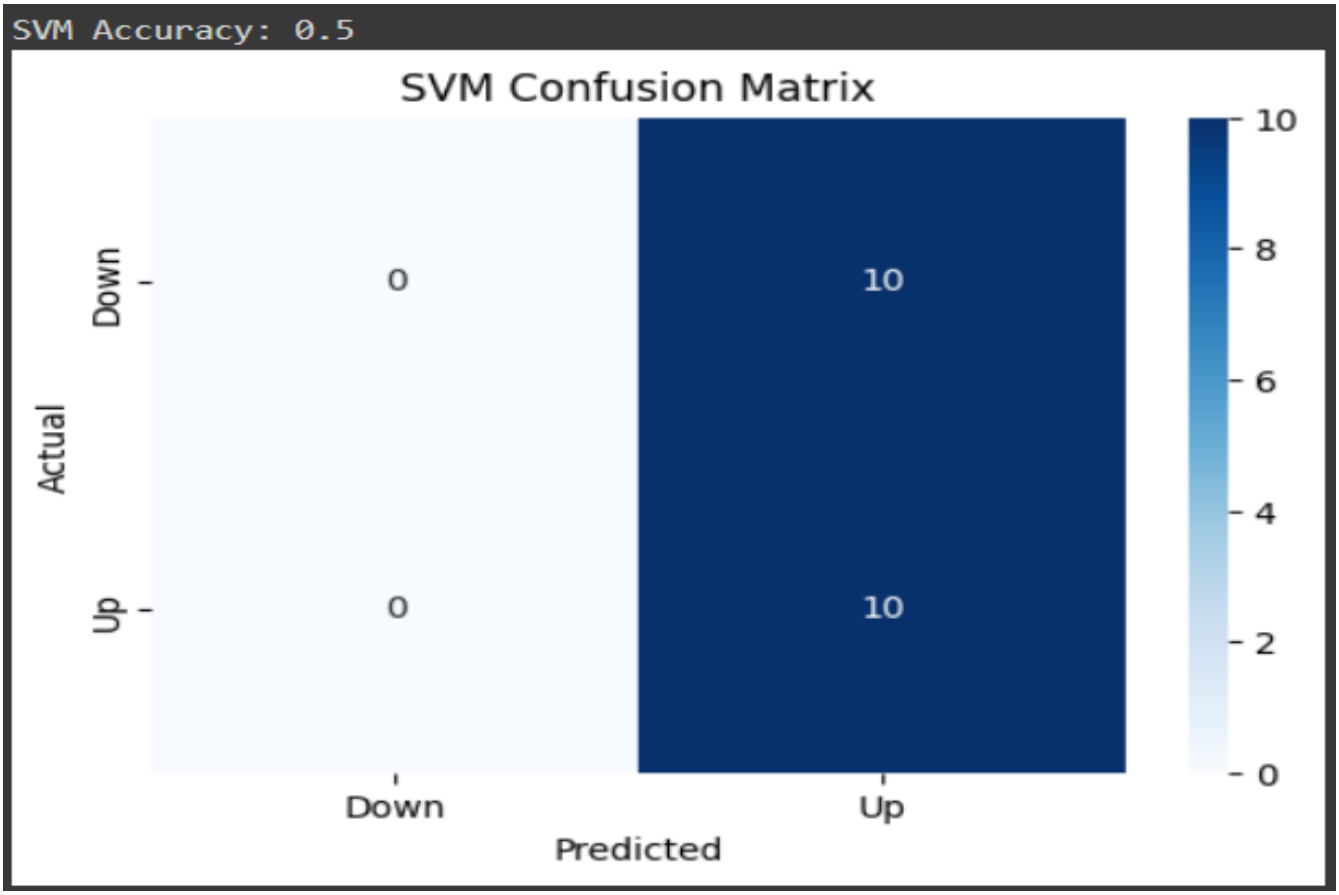
Model Evaluation

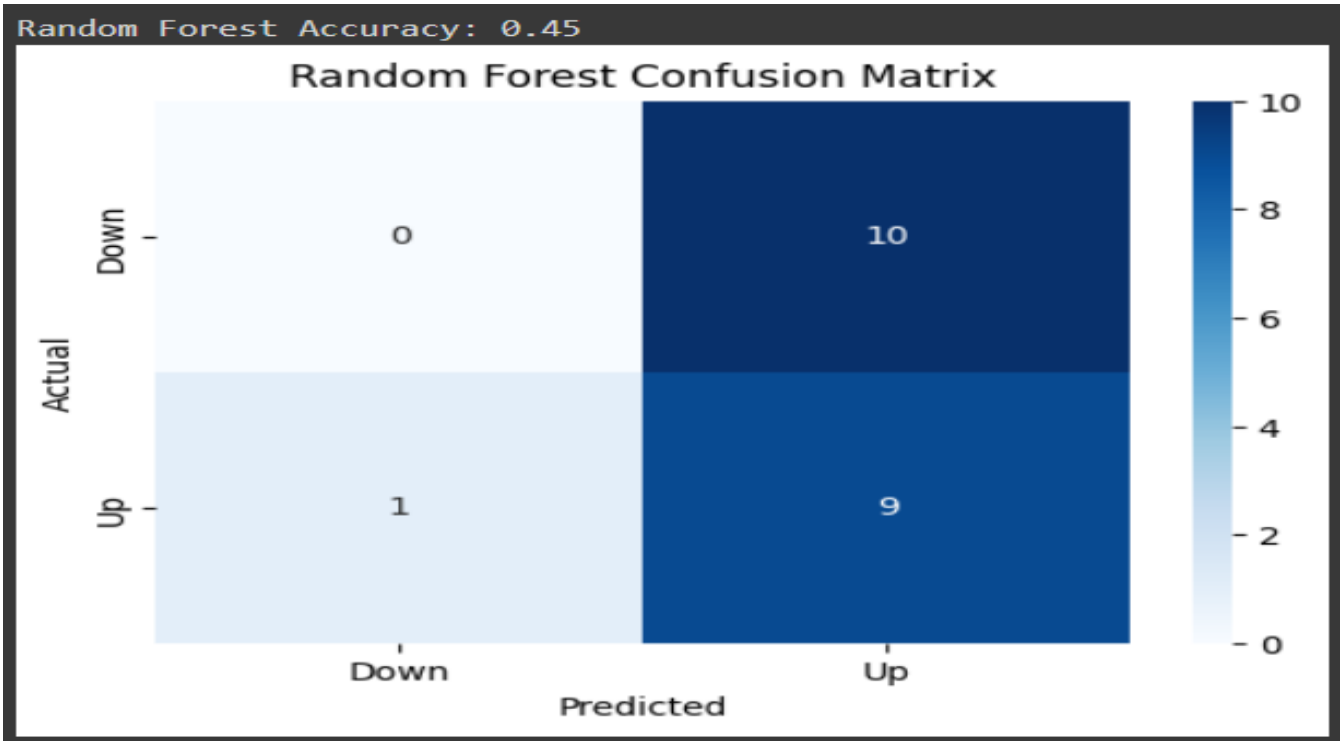
Models were evaluated using metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and classification accuracy for the directional movement prediction.

Results:




LSTM MAE: 7.4843756686316585
LSTM RMSE: 8.907529583180562
LSTM: Actual vs Predicted Stock Prices
Actual Prices
Predicted Prices
Stock Price
115
110
105
100
95
90
85
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5


SVM Accuracy: 0.5
SVM Confusion Matrix
Actual
Down
Up
0
10
0
10
Down
Up
Predicted
10
8
6
4
2
0



Discussion:

The analysis revealed that news and social media sentiments significantly influence stock market trends. Positive sentiments generally correlate with rising stock prices, while negative sentiments often precede declines. Among the models tested, the LSTM network demonstrated superior performance in capturing temporal dependencies and providing accurate predictions.

The findings highlight the potential of ML models to enhance traditional financial analysis by incorporating real-time sentiment data. However, the variability in social media data quality and the complexity of financial markets pose challenges that require ongoing refinement of models and methodologies.

Conclusion:

This study confirms the significant impact of news and social media on stock market trends and the utility of machine learning in analyzing these effects. By leveraging NLP and advanced ML models, investors and analysts can gain a deeper understanding of market sentiments and make more informed decisions. Future research should focus on improving data quality and exploring the integration of additional data sources to enhance prediction accuracy.

Список литературы:

1. Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2 (1), 1-8.

2. Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. Journal of Finance, 62 (3), 1139-1168.

3. Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. Journal of Finance, 66 (1), 35-65.

4. Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., & Ngo, D. C. L. (2014). Text mining for market prediction: A systematic review. Expert Systems with Applications, 41 (16), 7653-7670.

5. Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting stock market indicators through Twitter “I hope it is not as bad as I fear”. Procedia-Social and Behavioral Sciences, 26, 55-62.