Literature Review

Bhandari, Hum Nath, et al. “Predicting stock market index using LSTM.” *Machine Learning with Applications*, vol. 9, Sept. 2022, p. 100320, https://doi.org/10.1016/j.mlwa.2022.100320.

This study explores the use of Long Short-Term Memory (LSTM) neural networks to predict the next-day closing price of the S&P 500 index. The authors construct a model using nine well-balanced input variables, including fundamental market data, macroeconomic indicators, and technical indicators. They compare the performance of single-layer and multilayer LSTM models using metrics like Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the Correlation Coefficient (R).

The results indicate that the single-layer LSTM model, particularly with 150 neurons, outperforms multilayer models in terms of prediction accuracy and simplicity. The study demonstrates the LSTM's ability to capture complex, volatile stock market behavior more effectively than traditional time series models.

In this study, the authors use the S&P 500 index as the primary dataset for predicting stock price movements. The dataset spans a 15-year period from 2006 to 2020, which includes two significant bear markets: the 2008 financial crisis and the COVID-19 pandemic in 2020. This broad time frame was chosen to capture the market's behavior in both bullish and bearish conditions, enhancing the robustness of the predictive model.

The authors also acknowledge potential limitations, such as overfitting in multilayer models, and propose future improvements like incorporating unstructured data (e.g., social media sentiment) and exploring hybrid models to enhance prediction accuracy.

In conclusion, the study highlights the potential of single-layer LSTM models in predicting stock prices while maintaining a balance between complexity and accuracy, making it useful for investors and researchers.

Subasi, Abdulhamit, et al. "Stock market prediction using machine learning." Procedia Computer Science 194 (2021): 173-179.

Stock market prediction is complex due to the numerous factors influencing prices, such as investor sentiment, macroeconomic variables, and political events. This paper compares several machine learning classifiers to evaluate their predictive accuracy across major indices like NASDAQ, NYSE, Nikkei, and FTSE. By testing these models on both normal and leaked datasets, the study assesses how data integrity affects performance. The comparison highlights each algorithm's strengths in processing dynamic, noisy data, and offers insight into how they handle market volatility and random fluctuations typical of stock prices.

Leaked datasets significantly enhance performance, with Random Forest and Bagging classifiers achieving the highest accuracies (93%). Without leaked data, Support Vector Machines (SVM) often performed better, particularly for NASDAQ and NYSE. The findings highlight the superior predictive power of ensemble methods like Random Forest and Bagging when trained on more complete datasets, emphasizing the influence of data quality on stock prediction accuracy.

Lu, Minrong, and Xuerong Xu. "TRNN: An efficient time-series recurrent neural network for stock price prediction." Information Sciences 657 (2024): 119951.

https://www.sciencedirect.com/science/article/pii/S0020025523015360

Stock price prediction has long been a focal point in financial technology and machine learning research. Traditional approaches have employed neural network models such as Back Propagation (BP), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM). While these models have shown promise, they face challenges in training efficiency and accuracy. The importance of data preprocessing in big data analysis, particularly for prediction tasks, has been well-established, with time series-based processing methods highlighted as particularly effective.

The proposed Time-series Recurrent Neural Network (TRNN) model addresses several key areas in stock price prediction. It introduces a novel approach to time series data processing using sliding windows, which builds on existing literature on time series analysis in finance. The model's extraction of trends and turning points based on financial market features addresses a gap in existing models that often struggle with capturing market-specific characteristics. Furthermore, the data compression achieved through this method potentially addresses issues of computational efficiency noted in previous studies.

A significant innovation in the TRNN model is the upgrade of the price-volume relationship from one dimension to two dimensions based on RNN. This builds on existing research highlighting the importance of trading volume in stock price prediction. The comparative analysis of the TRNN model with RNN and LSTM models in terms of efficiency and accuracy provides valuable insights into the relative performance of these approaches, addressing calls in the literature for such comparisons.

The discussion of expanded schemes and the extendability of the time-series compression and TRNN model to other fields opens up interesting avenues for future research. This aligns with calls in the literature for more generalizable models in financial prediction. Overall, the TRNN model appears to address several gaps in the existing literature, particularly in terms of capturing market-specific features and improving computational efficiency in stock price prediction.