Machine Learning-Based Stock Market Price Prediction

By Pavan Reddy Narra

Abstract

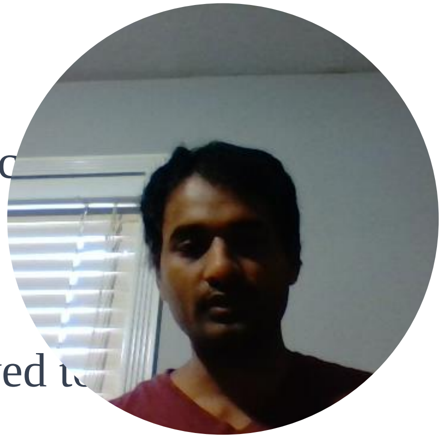
- Stock market price prediction is a challenging and complex task due to its dynamic environment

- The price of a stock depends on lots of factors

- Predicting the movement of the stock by using Machine learning may help the investors to take correct decisions in the right direction and mitigate the losses

Problem Statement

- Difficulty of predicting stock prices due to their complex, nonlinear nature influenced by various factors
- previous research has explored different approaches, including time series forecasting and deep learning, with varying degrees of success
- Some studies focused on technical analysis, while others combined technical and fundamental analysis.
- proposed approach suggests using a recent state-of-the-art transformer architecture that combines technical and fundamental analysis
- The primary objective is to assess whether this approach can achieve higher prediction accuracy c earlier models explored in the referenced research papers.
- Initially Block recurrent transformers was intend to use but due lack of implementation of it moved t recent state of algorithm TFT



Research Questions & Objectives

- What are the relevant data sources and features in order to extract the dynamic nature of the market?
- What are state of art algorithms to trace out the nonlinear relationship?
- What kind of evaluation metrics to use?
- To understand the correlation between the sentimental analysis and price changes
- To understand the patterns from historic data
- To compare the accuracy with other state of art algorithms
- To evaluate the performance of the model
- To integrate different sources of data to understand the dynamic variation of the markets

Scope of the Study

- The study is narrowing the focus to a select group of U.S. stocks.

- This study adopts a targeted approach by selecting specific sectors within the U.S. stock market. Within each sector, a diverse range of symbols has been selected

- The need to obtain a proper overview of market behavior is the driving force behind selecting various sectors and symbols in them.

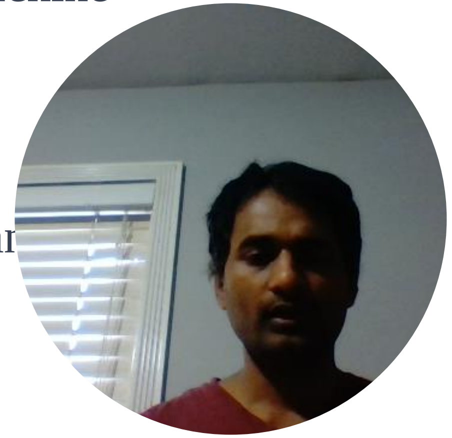
Literature Review

- RNNs, deep learning, frequency decomposition, self-attention transformers, hybrid information mixing modules

- Autoencoders, specifically deep autoencoders, are explored for feature extraction and denoising in stock market data

- RNNs, LSTMs, and GRUs for stock price prediction. It discusses their advantages, such as capturing complex patterns, and highlights the need for further research on long-term predictions and alternative machine learning techniques.

- Various evaluation metrics like MAE, RMSE, MAPE, and R2 are used to assess model performan



Literature Review

- Hybrid Models: Studies like CEGH (Liu et al., 2022) and HIMM (Choi et al., 2023) demonstrate improved performance compared to traditional models by combining different techniques.
- Transformer Models: Transformer-based models like TFT (Hu, 2021) and TAN (Qiuyue et al., 2022) effectively capture temporal dependencies, offering precise stock price predictions.
- Incorporating Multiple Data Sources: Including macroeconomic factors and sentiment analysis (Pahlawan et al., 2021, Parvatha et al., 2023) enhances prediction accuracy.
- Utilizing Time-Invariant Data: Integrating technical indicators (Lee et al., 2022) as time-invariant data improves prediction accuracy.
- Capturing Temporal Relationships: Transformer models learn complex temporal patterns and identify seasonal and weekly trends, enabling better predictions.

Research Gaps

- Inclusion of Quarterly Reports: Overlooked valuable financial data that can significantly impact stock prices.
- Identifying Long-Term and Short-Term Temporal Dependencies: Need to select models capable of recognizing both types of temporal connections.
- Incorporating Static Time-Invariant Features: Limited focus on attributes like industry sector and financial ratios that offer insights into long-term trends.

Methodology

- Initially thought of using Block Recursive transformers but due lack of libraries and code implementation switched to TFT

- TFT transformer is used for time series analysis

- Sentimental analysis is performed on the news articles

- The scores and sentimental category is appended to the historical price data and given as input to the model

Dataset

- Few challenges faced in collecting data
- Eodhistoricaldata.com API is used to collect the data which offers premium services
- Yahoo finance is used to collect the historical data
- Data is gathered combined and preprocessed

Why TFT model

- Capturing Temporal Dependencies TFT excels in capturing temporal dependencies in time series data.
- Stock market forecasting heavily relies on understanding these dependencies.
- TFT offers a comprehensive approach by incorporating time-variant and time-invariant factors.
- TFT's ability to handle diverse data sources and temporal relationships contributes to improved forecasting accuracy.

Why TFT model

- Recurrent Layers: Incorporating recurrent layers for local processing to capture short-term dependencies.
- Self-Attention Layers: Leveraging self-attention mechanisms to understand long-term dependencies.
- Gating Mechanisms: Regulating information flow using Gated Linear Units (GLU) and variable selection networks.
- It uses Temporal Fusion Decoder to identify nearby values and to maintain dataflow

Model Implementation

- The following features were given as input to the Model

  open,close,volume,high,low,sentimentalscore,compound_sentimental   score

- The model is implemented by using pytorch

- Timeseries data set is prepared with sequence length of 5 and max prediction length of 2

- Hyperparameters is tuned with optuna

Results and Analysis

- LSTM,GRU,BILSTM attention is used for comparison
- These models are chosen because those are widely used in the research papers in the literature review
- All the models got R2 of 99% and GRU having low MAE of 1.78 followed by LSTM, TFT, Attention Based BILSTM
- But when performed on individual stocks TFT model performance is better than other models

Conclusion and Future analysis

- Extend Data Period and Diversify Sources
- Incorporate Sentiment Analysis
- Multi-Model Approach
- Feature Engineering