Portfolio Optimization and Stock Returns Prediction Using Neural Networks and Supervised Models: Comparative Study of LSTM, CNN, RF, LR Models

Aigerim Mansurova

Pace University, Seidenberg School of CSIS

PACE UNIVERSITY

Abstract

Predicting stock returns using Deep Neural Network (DNN) and Supervised Learning models is in high demand nowadays because it has the potential to provide more accurate and reliable information compared to traditional methods. In this project, long short-term memory (LSTM) neural network, convolutional neural network (CNN), Random Forest (RF) and Linear Regression (LR) models are employed to predict returns of stocks listed in S&P 500. Then, mean-variance optimization model is applied to calculate optimal weights of stocks in a predication-based portfolio.

Research Questions

Which forecasting model performs the best in predicting the returns of stocks?

How can optimization technique be used to determine the optimal allocation of assets in a portfolio for maximum return and minimal risk?

Related Work

Several studies have explored the use of deep learning techniques for portfolio optimization and stock prediction in the financial market [Y. Ma, R. Han , W. Wang, 2020]. [Fischer, Krauss, 2018] used long short-term memory networks for financial market predictions, and [Wang et al., 2020] presented a deep learning approach for portfolio formation with preselection using long-term financial data.

Dataset

5-year daily historical stock prices (01.01.2018 – present day) of major four Sectors (Energy, Financials, IT, Health Care), listed in S&P 500, are collected from Yahoo Finance API from 01.01. 2018 to present day. For each stock, daily 'Open', 'Close', 'Low', 'High' prices and trading volume were used as the main input values in the dataset. Overall, the shape of the dataset is 1305 rows by 10[illegible]

stock_df

Ticker	ABT					ABBV		
Stock Info	Open	High	Low	Close	Volume	Open	High	Low
Date								
2018-01-02	58.200001	59.200001	57.820000	58.790001	10112800	97.139999	98.900002	96.75
2018-01-03	58.990002	59.020000	58.310001	58.919998	5683700	98.550003	100.099998	98.04
2018-01-04	59.500000	59.599998	58.759998	58.820000	6240000	100.070000	100.120003	98.34
2018-01-05	59.040001	59.090000	58.639999	58.990002	5836900	99.339996	101.199997	98.55
2018-01-08	58.849998	58.980000	58.529999	58.820000	5411500	101.279999	101.279999	98.18
...	...	...	...	...	...	...	...	

Methodology

Data Preprocessing

This project applies past 20 days' daily growth of open, close, high, low price and volumes, having a dimension of input for each prediction as 20x5, to predict next day's return. Growth is calculated by the formula:

where p(t) is current price,

p(t-1) – last day's price.

$$r_t = \frac{p_t - p_{t-1}}{p_{t-1}}$$

Extra high and low values are detected as follows:

$$d_i = \begin{cases} d_m + 5d_{mm} & if\ \ d_i \geq d_m + 5d_{mm}, \\ d_m - 5d_{mm} & if\ \ d_i \leq d_m - 5d_{mm}. \end{cases}$$

where dm is median of each feature (di); dmm - median absolute deviation. Then, data is standardized. Sliding window is used, i.e., the first 2 years data is training set, following 1 year is validation and data from 2022-01-01 to present day is testing set.

Modeling

Four different models (LSTM, CNN, RF, LR) are used to see the testing score and say which one performs best. This experiment applies mean-squared error and R-squared as model performance metrics.

Portfolio Optimization Model

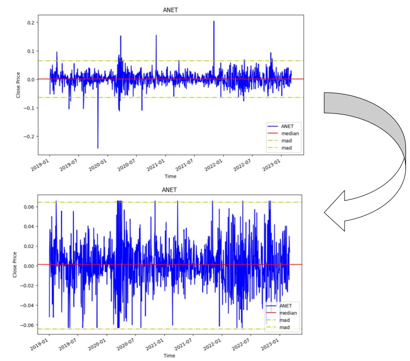
Based on the averages of predicted returns, the top 5 stocks from each sectors are selected to form diversified portfolio. Portfolio allocation is then performed using mean-variance optimization technique. Mean-variance method aims to find optimal weights to maximize the expected return of portfolio for a given level of risk aversion using following formula:

$$\begin{aligned} \min_{w}\ \bar{\sigma} &= w^T \Sigma w \\ s.t. & \\ & w^T \hat{\mu} = \bar{r} \\ & w^T 1 = 1 \end{aligned}$$

A higher value of risk aversion implies that the investor is more risk-averse and would require a higher expected return to compensate for the additional risk.

Results

Below graphs represents the stocks' prices before and after applying outliers_detection function:



After fitting given models, the mean and standard deviations of MSE and R^2 metrices are calculated and presented in the following table:

Metric	Mean		Std	
Score	MSE	R2	MSE	R2
Model				
LSTM	0.12736	0.880719	0.054511	0.027696
CNN	0.130962	0.8767	0.054733	0.029555
RF	0.15002	0.857822	0.058349	0.0311
LR	0.129761	0.87765	0.051735	0.027885

LSTM has the lowest mean of MSE and highest one of R^2 compared to the other models', and although their standard deviations are not the lowest, their difference is relatively small. On the other hand, Linear Regression's mean and standard deviations of model performance metrics shows a good result, which means supervised learning model can be also used as a predictor of returns. Nevertheless, LSTM is a superior model in stock returns prediction.

Thus, predicted returns by LSTM models are selected to construct a portfolio.

	Symbol	Average_Return	Name	Sector
0	WST	1.093681	West Pharmaceutical Services	Health Care
1	ANET	1.064967	Arista Networks	Information Technology
2	ANSS	1.024878	Ansys	Information Technology
3	NVDA	0.903886	Nvidia	Information Technology
4	CRM	0.617188	Salesforce	Information Technology
5	JNPR	0.353130	Juniper Networks	Information Technology
6	INTC	0.326263	Intel	Information Technology
7	FTNT	0.322509	Fortinet	Information Technology
8	ABBV	0.311089	AbbVie	Health Care

After setting risk aversion to 0.3 and applying mean_variance_optimization function the following result is obtained:

	Equal weights	Optimal weights
Expected annual return of portfolio	7.3%	17.52%

Conclusion and Future works

This project applies two DNN and two supervised learning models to predict returns of stocks listed in S&P 500 index. The best-performing model is chosen by comparing the mean and standard deviation of evaluation metrics (MSE and R^2). LSTM, which is a frequently used DNN model that has been shown to have better learning abilities than traditional ML models, has a mean MSE of 0.13 and a mean R^2 of 0.88. Portfolio is constructed by top five stocks, having the highest average returns, from each sectors. Mean-variance optimization technique is utilized to calculate the optimal stock weights allocation. Expected annual returns of portfolio is calculated using two different methods: optimal weights obtained from applying MVO technique and equal weights. The return from equal weights is 7.3%, while the return from optimal weights is 17.52%.

The performance of models could be improved by adding new features such as economic, financial indicators and applying NLP by considering news headlines. Additionally, there might be other optimization techniques other than mean-variance optimization model. Future research could expand the input features and utilize superior risk metrics to construct prediction-based portfolio models, improving their out-of-sample performance.

References

- Y. Ma, R. Han , W. Wang*: Prediction-Based Portfolio Optimization Models Using Deep Neural Networks,* IEEE Xplore Digital Library: 10.1109/ACCESS.2020.3003819
- Van-Dai Ta, Chuan-Ming Liu, Direselign Addis Tadesse: *Portfolio Optimization-Based Stock Prediction Using Long-Short Term Memory Network in Quantitative Trading,* Applied Science
- T. Fischer and C. Krauss*: Deep learning with long short-term memory networks for financial market predictions,'* Eur. J. Oper. Res., vol. 270, no. 2, pp. 654–669, Oct. 2018.
- W. Wang, W. Li, N. Zhang, K. Liu: *Portfolio formation with preselection using deep learning from long-term financial data*, Expert Syst. Appl., vol. 143, Apr. 2020, Art. no. 113042.
- A. J. P. Samarawickrama, T. G. I. Fernando: *A recurrent neural*
- *network approach in predicting daily stock prices an application to the Sri Lankan stock market*, in Proc. IEEE Int. Conf. Ind. Inf. Syst. (ICIIS), Dec. 2017, pp. 1–6.