Predicting Stock Prices Using Different Machine Learning and Deep Learning Models

GC-511

Presenters: Afnan Crystal, Rohith Sundar, and Hrithik Singh | Advisor: Dr. Mahmut Karakaya

College of Computing and Software Engineering, Kennesaw State University

Introduction

- Our project focuses on predicting daily closing prices and stock movements of Amazon, a dynamic corporation known for unpredictable stock prices influenced by various factors
- We employ a comprehensive approach, comparing the performance of Linear Regression, Support Vector Machine (SVM), and Multi-Layered Perceptron (MLP) models on financial time series data from January 2, 2005, to August 21, 2019
- Our goal is to provide accurate forecasts and our project outlines the methodologies and techniques used for stock price forecasting
- **Preliminary Results:** Pilot study shows promising results; we achieved an overall accuracy of 93% on five categories of food.

Method Overview

- **Data Preprocessing:** Firstly, we prepare and clean the dataset and target column to create the prediction target which is the closing price for the following day
- **Model Selection:** We use these three diverse machine learning & deep learning models for our prediction tasks
  i. Linear Regression (LR)
  ii. Support Vector Machine (SVM)
  iii. Multi-Layered Perceptron (MLP)
- **Training the Models:** The model optimizes SVM through grid search for hyperparameter tuning, requires no tuning for Linear Regression, and implements early stopping in a two-layer MLP to prevent overfitting and enhance generalization
- **Predictions and Evaluation:** Model performance is assessed using standard regression metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R2) score

Model Training/Dataset

We generated a "Target" column by shifting the stock closing prices by one time step and applied .MinMax scaling, ensuring values are within the [0, 1] range

We split the dataset to use 80% of the data for training and 20% of the data for testing using a train-test split with the shuffle parameter set to False to preserve the temporal order of the data

These preprocessing steps are essential for creating a well-structured, temporally aligned dataset, ensuring effective model training and accurate predictions in the domain of stock price forecasting

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Date	Open	High	Low	Close	Volume	Dividends	Stock Splits	Gain	Loss	Avg_Gain	Avg_Loss	RSI
2	1/3/2005	44.95	45.44	44.21	44.52	1E+07	0	0	0.23	0	0.57214	0.23714	70.6973
3	1/4/2005	42.67	43.26	41.5	42.14	1.9E+07	0	0	0	2.38	0.52857	0.40714	56.4885
4	1/5/2005	41.57	42.76	41.56	41.77	8354200	0	0	0	0.37	0.48857	0.43357	52.9822
5	1/6/2005	41.81	42.25	40.9	41.05	8700900	0	0	0	0.72	0.48857	0.41714	53.9432
6	1/7/2005	41.38	42.69	41.16	42.32	9836600	0	0	1.27	0	0.57929	0.41429	58.3034
7	########	41.94	42.96	41.71	41.84	7347900	0	0	0	0.48	0.57929	0.37286	60.8402
8	########	41.4	42.16	41.01	41.64	7920300	0	0	0	0.2	0.54786	0.38714	58.5943
9	########	41.49	42.48	40.82	42.3	8072300	0	0	0.66	0	0.58571	0.38714	60.2056
10	########	42.45	44.58	42.34	42.6	1.7E+07	0	0	0.3	0	0.60714	0.345	63.7659
11	########	42.9	44.66	42.36	44.55	1.3E+07	0	0	1.95	0	0.50929	0.345	59.6154
12	########	44.29	45	44.15	44.58	7131600	0	0	0.03	0	0.34143	0.345	49.7399
13	########	44.47	44.74	43.43	43.96	7396700	0	0	0	0.62	0.32786	0.38929	45.7171
14	########	41.71	43.96	41.64	42.36	1.3E+07	0	0	0	1.6	0.31714	0.50357	38.6423
15	########	41.99	42.55	41.02	41.16	1.1E+07	0	0	0	1.2	0.31714	0.54071	36.9692
16	########	41.44	41.95	40.31	40.38	9695900	0	0	0	0.78	0.30071	0.59643	33.5191
17	########	40.64	41.44	40.45	40.94	9027900	0	0	0.56	0	0.34071	0.42643	44.4134
18	########	41.18	41.8	40.26	41.34	7714900	0	0	0.4	0	0.36929	0.4	48.0037
19	########	41.27	42.4	41.14	42.31	8038400	0	0	0.97	0	0.43857	0.34857	55.7169
20	########	42.09	43	41.53	42.22	6152400	0	0	0	0.09	0.34786	0.355	49.4919
21	########	42.44	43.42	42.43	43.22	6320300	0	0	1	0	0.41929	0.32071	56.6602
22	2/1/2005	43	43.27	42.38	42.48	6761500	0	0	0	0.74	0.41929	0.35929	53.8532
23	2/2/2005	43.05	43.54	40.86	41.88	2.6E+07	0	0	0	0.6	0.37214	0.40214	48.0627
24	2/3/2005	34.9	35.89	34.53	35.75	6.1E+07	0	0	0	6.13	0.35071	0.84	29.4541
25	2/4/2005	35.65	35.98	35.19	35.72	1.7E+07	0	0	0	0.03	0.21143	0.84214	20.0678
26	2/7/2005	35.82	35.84	35.55	35.69	7112400	0	0	0	0.03	0.20929	0.84429	19.8644
27	2/8/2005	35.54	36.52	35.53	36.3	9602500	0	0	0.61	0	0.25286	0.8	24.0163
28	2/9/2005	36.45	36.8	35.87	35.89	7938800	0	0	0	0.41	0.25286	0.715	26.1255
29	########	36.07	36.11	35.53	35.78	7982400	0	0	0	0.11	0.25286	0.63714	28.4109

Algorithms Deployed

- **Linear Regression :** In this context, it is employed to predict the next day's closing stock price for Amazon. Linear Regression provides clear insights into the relationship between the input features (such as 'Close' and 'Google_Trends') and the target variable ('Target' column representing the next day's closing price). The coefficients in the linear equation signify the impact of each feature on the prediction.
- **Support Vector Machine:** SVM is employed as a regression model to forecast the next day's closing stock price for Amazon. Hyperparameters for the SVM model are optimized using GridSearchCV where the hyperparameters include 'C' (regularization parameter), 'kernel' (type of kernel function), and 'gamma' (kernel coefficient). SVM complements Linear Regression in this project, offering a more sophisticated approach to capture non-linear relationships, potentially improving predictive accuracy. The grid search optimizes the SVM model for the data, enhancing its performance in predicting Amazon's stock prices.
- **Multi-Layered Perceptron (MLP):** MLP is constructed and trained to forecast the next day's closing stock price for Amazon. In the project a neural network model is constructed using Keras. It consists of two hidden layers with 50 units and ReLU activation functions. The model is trained using Mean Squared Error (MSE) as the loss function and the Adam optimizer where we used Early stopping to prevent overfitting.

Results

The SVM model has the highest MSE, indicating that on average, the squares of the errors are the largest among the three models. However, its MAE is competitive, suggesting that when considering absolute errors (without squaring them), it performs better.

The Linear Regression model shows the lowest MSE and MAE, indicating it has the closest predictions to the actual values on average. Its R^2 value is also the highest, suggesting that it explains the variance of the dependent variable best out of the three models.

The MLP model, while having the highest MAE, still shows a relatively high R^2 value, indicating a good fit to the data, but perhaps with some outliers or variability that it doesn't handle as well as Linear Regression.

Model	MSE	MAE	R2
SVM	7928.7666	80.927534	0.953357
Linear Regression	724.32148	17.444968	0.995739
MLP	905.7408	21.810191	0.994672

Conclusion

Our project, which aims to predict Amazon's daily closing stock prices using various machine learning and deep learning models, has revealed that the Linear Regression model outperforms Support Vector Machine and Multi-Layer Perceptron.

When used on financial time series data of Amazon, showing an unprecedented accuracy with a high R-squared value of 0.9957, while achieving the lowest Mean Squared Error (MSE) and Mean Absolute Error (MAE).

These preliminary findings underscore the significance of model selection in stock price forecasting and provide valuable insights for investors and financial institutions seeking reliable tools for decision-making.