Introduction

Title: Analysis of Stock Prices Using Machine Learning and Deep Learning Techniques

Opening Paragraph:

The ever-evolving stock market presents a multitude of challenges and opportunities for investors and analysts. Navigating through its unpredictability requires not only keen insight but also the aid of advanced analytical tools. This report delves into the realm of machine learning and deep learning, aiming to harness their predictive power in the context of stock market analysis.

Background:

The stock market's inherent volatility and complexity make accurate forecasting a highly sought-after yet challenging goal. Traditional methods, while valuable, often fall short in addressing the rapid changes and vast data involved in stock market analysis. The advent of machine learning and deep learning opens new frontiers in this domain, offering sophisticated techniques capable of handling large datasets and identifying complex patterns.

Purpose:

This project aims to explore and evaluate the effectiveness of various machine learning and deep learning models in predicting stock prices. By analyzing historical data of key stocks from different sectors, particularly banking and technology, we seek to understand how these models can enhance prediction accuracy and offer valuable insights for stock market participants.

Scope:

The report covers a range of machine learning models, including Linear Regression, Ridge, Lasso, and SGD, alongside classification models like Decision Trees, Random Forest, and Deep Neural Networks. The study encompasses data collection, preprocessing, model implementation, performance evaluation, and an interpretation of the results, providing a holistic view of the potential and limitations of these technologies in stock market analysis.

Methods

Data Collection and Preprocessing:

The project utilizes the 'yfinance' library for collecting historical stock data, focusing on key financial stocks from the banking and technology sectors. This data includes vital metrics like closing prices, volumes, and market trends. Preprocessing steps involve cleaning the data, handling missing values, and normalizing the dataset to ensure uniformity for analysis. Additionally, the data is split into training and testing sets to validate the models' predictions.

Model Implementation:

The study employs a range of machine learning models:

- Regression Models: Linear Regression, Ridge, Lasso, and SGD. These models are used for predicting continuous outcomes, such as stock prices.
- Classification Models: Decision Tree, Random Forest, and KNN. These models predict categorical outcomes like the direction of stock price movement.
- Deep Learning Model: A Deep Neural Network (DNN) is implemented to explore advanced predictive capabilities. The DNN architecture is designed with multiple layers to capture complex patterns in the data.

Model Evaluation:

Each model's performance is evaluated using relevant metrics. For regression models, the Mean Square Error (MSE) is used to measure the performance of the stock price prediction models. For classification models, Accuracy is the criteria for comparing classification of direction of movement.

Results

Analysis of Regression Models:

The regression models, including Linear, Ridge, Lasso, and SGD, demonstrated varying levels of accuracy in predicting stock prices. The Mean Square Error (MSE) was used as the primary metric for evaluation. Certain models performed better with specific stocks, indicating the influence of stock characteristics on model performance.

Table 1 shows the results of the regression models fitted to the sectores and their related stocks.

Table 1. Mean Squared Error for Different Stocks Close Price Forecasting with different Regression models

	Sector											
	Banking			Technology			Manufacturing			Communication		
Model	JPMorgan Chase (JPM)	Bank of America (BAC)	Wells Fargo (WFC)	Apple (AAPL)	Microsoft (MSFT)	Alphabet Google (GOOGL)	General Electric(GE)	3M(MMM)	Honeywell(HON)	AT&T(T)	Verizon(VZ)	Comcast(CMCSA)
Linear	**1.4946**	**0.4331**	0.6522	**1.4129**	**2.6793**	**1.1246**	**1.5289**	**1.84**	**1.7580**	0.1904	**0.3882**	**0.4589**
Ridge	1.5033	0.4384	**0.6514**	1.4542	2.7234	1.1514	1.5433	1.8647	1.7903	0.1911	0.3899	0.4580
LASSO	1.8834	1.1069	1.2030	1.8600	3.0417	1.5229	1.8773	2.1819	2.121	1.0858	1.1188	1.144
SGD	1.5476	1.5476	0.6814	1.5266	2.7916	1.2046	1.6335	1.9332	1.8603	**0.1979**	0.4151	0. 4817
Decision Tree	2.1280	0.5657	0.8830	2.1030	3.7063	1.667	2.2147	2.5729	2.6241	0.2927	0.6114	0.6761
Random Forest	1.6281	0.4607	0.6987	1.5901	2.8644	1.2635	1.6728	2.023	1.9449	0.2215	0.4523	0.4994
DNN	1.7399	0.5380	0.6856	1.4912	2.8842	1.7840	1.5441	1.9095	2.2133	0.2652	4.8370	0.5216
Average for each stock	1.6530	0.4910	0.6976	1.6752	2.9868	1.2616	1.7163	2.0464	2.0445	0.3493	1.1732	0.6056
Average for each sector	0.9472			1.9745			1.9357			0.7093		

Classification Model Findings:

The Decision Tree, Random Forest, and KNN models were evaluated based on their accuracy in predicting the direction of stock price movement. Results showed a significant variation in performance across different models and stocks, suggesting that some models are more suited to specific market conditions than others. The result of classification models are summarized in Table 2.

Table 2. Accuracy for Different Stocks Movement Direction Forecasting with different Classification models (%)

Model	Sector												
	Banking			Technology			Manufacturing			Communication			Average Accuracy for each model
	JPMorgan Chase (JPM)	Bank of America (BAC)	Wells Fargo (WFC)	Apple (AAPL)	Microsoft (MSFT)	Alphabet Google (GOOGL)	GE	MMM	HON	T	VZ	CMCSA	
K Nearest Neighbor	0.4821	**0.5135**	0.4950	0.5007	0.4921	0.4964	0.5092	0.5121	0.5363	0.4964	0.5078	0.4893	0.5025
Decision Tree Classifier	0.4880	0.4781	0.5119	0.5	**0.4960**	0.5337	**0.5198**	0.5460	**0.5575**	0.4880	0.4801	0.5337	**0.5194**
DNN	**0.5021**	0.4793	**0.5235**	**0.55066**	0.4935	0.4978	0.5121	**0.5606**	0.5534	**0.5278**	**0.5349**	**0.5492**	0.5078
Average accuracy for each stock	0.4907	0.4903	0.5101	0.5171	0.4938	0.5093	0.5146	0.5395	0.5490	0.5040	0.5076	0.5240	
Average accuracy for each sector	0.4970			0.5067			**0.5343**			0.5118			

Table 3.Comparing Models Performance

Model	Selected as Best Model
Regression	
Linear	11
Ridge	1
LASSO	0
SGD	0
Random Forest	0
Classification	
K Nearest Neighbor	1
Decision Tree Classifier	4
DNN	7

Based on the results in Table 3, for regression task, the linear regression task out performs all other models by lowest MSE. By implementing other complicated models like Geometric Brownian Motion, it is possible to reduce MSE and fit better models.

Regrading Classification task, DNN superior performance is shown highest accuracy in 7 out of 12 stock.

Deep Neural Network (DNN) Performance:

The Deep Neural Network model, designed with multiple layers, showed promising results, particularly in capturing complex patterns in stock price movements. Its performance, in terms of both accuracy and reliability, was noteworthy compared to traditional machine learning models, especially in volatile sectors like technology.

Sector-Specific Insights:

A comparative analysis across sectors revealed that models' performance varied significantly between the banking and technology sectors. This emphasizes the importance of choosing the right model based on sector-specific characteristics.

In regression analysis with four sectors, the communication sector has the lowest MSE, which make sense. Because the communication is not a industry with high fluctuations. Among all the stocks studied is this project, the AT & T stock from communication sector has the lowest MSE.

In Classification analysis, highest average classification accuracy belongs to the Manufacturing sector which can be an indicator of the low variation in that sector. Also, the highest average classification accuracy among the all 12 stocks belongs to HON stock from the Manufacturing sector.

Conclusion

Summary of Findings:

The project successfully demonstrated the application of machine learning and deep learning models in predicting stock prices. Results indicated that different models have varying degrees of effectiveness, with deep learning models, particularly Deep Neural Networks, showing superior performance in classification task.

Final Thoughts:

This study underscores the importance of selecting the right model based on the specific characteristics of the stock and sector. As machine learning and deep learning continue to evolve, their role in financial analysis is likely to become increasingly significant.