Stock Price Prediction: A Hybrid Approach Using Machine Learning and Deep Learning

# Abstract

Predicting stock prices is challenging. It’s like trying to guess the weather—one minute it’s sunny, and the next it’s raining. But what if we had a tool that could help us predict it more accurately?  
  
In this study, we used machine learning (ML) and deep learning (DL) to predict stock prices. We tested different models, such as Random Forest, Decision Tree, Gradient Boosting, and Long Short-Term Memory (LSTM). We also combined them into one hybrid model. The results show that using ML and DL together improves prediction accuracy. This method could help investors and analysts make more reliable decisions.

# Introduction

The stock market is unpredictable. One minute, prices rise; the next, they fall. It’s like trying to catch a basketball in the dark. Predicting where they go next? It’s tough. Like guessing the next twist on a rollercoaster ride, it’s nearly impossible.  
  
I recently asked GPT to predict the stock price of a company. Based on the latest news, it wasn’t perfect, but it was more accurate than my own guess. That made me think—what if we used something more powerful, like machine learning (ML) and deep learning (DL)?  
  
These tools are like detectives for data. They spot clues. They look for patterns. Then, they use these patterns to predict what’s coming next. This is why we turned to ML and DL for stock prediction. They have the power to analyze huge amounts of data and help make more accurate predictions.  
  
Why use ML and DL for stock predictions?  
  
It’s like solving a puzzle with missing pieces. ML and DL fill in those gaps. They learn from the past, just like a weather app predicting rain. They can handle messy data, like finding the clean sock in a laundry basket.  
  
In this study, we tested several models:  
  
- Random Forest: A group of trees working together to make decisions. It’s like a team of experts coming together to solve a problem.  
- Decision Tree: A single tree making choices step by step. It’s like following a simple recipe.  
- Gradient Boosting: Trees that learn from their mistakes. It’s like revising your work after seeing where you went wrong.  
- Long Short-Term Memory (LSTM): A neural network that remembers past information. It’s like having a memory that helps predict what’s coming next.  
  
Each model has its strengths. But when combined, it’s like making the perfect sandwich. ML gives its structure, while DL adds the flavor. Together, they produce better predictions.

# 2. Literature Review

2.1 Traditional Models  
Traditional models like ARIMA are useful but limited. They’re like using an old map to find a new place, helpful, but not always accurate. The stock market is more complicated than these models can handle.  
  
2.2 Machine Learning Models  
Machine learning models, like Random Forest and Gradient Boosting, are good at finding patterns in large datasets. It’s like teaching a dog new tricks, the more you show it, the better it gets. But these models don’t do as well with time-based data.  
  
2.3 Deep Learning Models  
Deep learning models, especially LSTMs, are great for handling time-based data. They work like a memory that learns from past events. However, they require careful management. It’s like maintaining a high-performance car. Without the right tuning, it won’t run as smoothly.  
  
2.4 Hybrid Approaches  
Combining ML and DL is like mixing the best of both worlds. ML models are stable and provide structure, while DL models excel at understanding time-based patterns. Together, they create a more accurate prediction model.

# 3. Methodology

3.1 Data Source and Characteristics  
We used stock data from three companies: ADANI PORTS, TATA GLOBAL, and ZOMATO. The dataset includes:  
  
- Date: The trading day.  
- Close Price: The stock price at the end of the day (this is what we want to predict).  
- Volume and Trades: The number of shares traded each day.

3.2 Preprocessing Steps  
Handling Missing Values: Missing data was filled in using linear interpolation. Think of it as filling in the gaps in a story.

3.3 Normalization We scaled the data to ensure everything was on the same level. It’s like adjusting the volume so it sounds just right.  
Sequence Preparation: For LSTM models, we created 60-day windows to predict the next day’s price. It’s like looking at the last 60 days to predict what will happen tomorrow.

# 4. Machine Learning Models

**4.1 Random Forest**

Random Forest is like a team of decision trees. Each tree looks at the data and makes its own prediction. Then, the team combines the predictions. It’s like a group of experts working together to make the best decision.

**4.2 Decision Tree**

A Decision Tree is simpler. It’s like following a flowchart where each decision leads you to the next step. It’s easy to understand, but sometimes it can miss the bigger picture if the data is too complex.

**4.3 Gradient Boosting**

Gradient Boosting builds trees one by one. Each new tree corrects the mistakes of the previous one. It’s like editing your work, fixing one mistake at a time until it’s perfect.

## 5. Deep Learning Models

LSTMs are designed to remember what happened in the past. They have three gates:

* **Forget Gate:** Decides what information to discard.
* **Input Gate:** Decides what new information to keep.
* **Output Gate:** Decides what to predict.

It’s like a high-performance sports car that remembers the best routes. But just like a car needs maintenance, LSTMs need proper management to avoid going off track.

**5.2 Model Compilation**

We set up the LSTM model with:

* **Loss Function:** Mean Squared Error (MSE), which helps us measure prediction accuracy.
* **Optimizer:** Adam, which automatically adjusts the learning rate, much like setting the perfect speed for a car.

## 6. Hybrid Approach

**6.1 Model Integration**

We combined predictions from Random Forest, Decision Tree, Gradient Boosting, and LSTM. It’s like having a team of experts, each with different skills, coming together to solve a problem.

**6.2 Weighted Averaging**

We gave each model a weight based on how well it performed. The stronger models had more influence, just like letting the most experienced expert speak more often.

# 7. Results

**7.1 Machine Learning Performance**

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **MSE** | **MAE** | **R²** |
| **Random Forest** | **0.019** | **0.092** | **0.87** |
| **Decision Tree** | **0.023** | **0.101** | **0.83** |
| **Gradient Boosting** | **0.018** | **0.089** | **0.88** |

**7.2 LSTM Performance**

* **MSE:** 0.015
* **MAE:** 0.083
* **R² Score:** 0.91

**7.3 Hybrid Model Performance**

* **MSE:** 0.012
* **MAE:** 0.078
* **R² Score:** 0.94

## 8. Discussion

**8.1 Key Observations**

* Random Forest performed best among the ML models.
* LSTM worked well for capturing patterns over time.
* The hybrid model performed the best overall.

**8.2 Error Analysis**

* ML models struggled with sudden changes in the market.
* LSTM sometimes overfitted, but regularization helped address that.

## 9. Limitations and Future Work

**9.1 Limitations**

* We only used one year of data.
* We only looked at closing prices.

**9.2 Future Directions**

* We plan to add more features, like economic indicators (GDP, inflation).
* We will test longer datasets to improve predictions.

**10. Conclusion**

Combining ML and DL is like using the best tools for the job. ML provides stability, while LSTMs capture time-based patterns. This hybrid approach improves the accuracy of stock price predictions and can help investors make better decisions.