JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

HYBRID MACHINE LEARNING ALGORITHMS BASED ON STOCK PRICE PREDICTIONS

[1]Ch Harshavardhan, [2]B Siddhartha, [3]T Rahul, [4]N Pavan Kumar, [5]A Ravi Kumar
[1,2,3,4,] IV B.Tech Students, [5]Assistant Professor
Dept. of Information Technology, Sreenidhi Institute of Science and Technology(Autonomous), Hyderabad, India

Abstract: A stock market is an open market or public market for buying and selling shares for publicly listed companies. Different types of stocks are known for equities and ownership in the company. Mediator for buying and selling of shares monitored by the stock exchange. The share market hugely impacts customers, businessmen, and financial institutions. In present days advanced technologies are used for data classification and analysis purposes. Using technical models in financial institutions for price prediction, investment opportunities, investments, and portfolio optimization, etc. Our present research uses machine learning models for prediction for financial trend analysis. It shows a comprehensive review of practical applications of machine learning models. A Mean Squared Error (MSE) value of 77.58 for an ensemble model combining Linear Regression, Ridge Regression, and Elastic Net Regression suggests that the model's predictions exhibit a moderate level of error when compared to the true target values.

Index Terms: Machine Learning models, Stock Prices, Predictions, Accuracy, MSE

I. INTRODUCTION

In public listed companies sell the shares to customers by stock exchange in stock market. These stocks are also known as ownership in the company [1]. The stock exchange monitors all the shareholders for buying and selling of shares. Stock markets connect with a huge amount of data, data analysis is the toughest context for handling [2]. So advanced technologies are used to analyze the data. Machine learning models support finding the future prediction of company stocks, financial assets for traders, and customer share data [3]. Monitoring significant profits/loss of shares of customers predicting whole data and stock performance is also hard to monitor for analysis. So many factors impact on prediction of future data analysis [4]. These factors integrate to make share prices volatile and dynamic. These models work on stock price prediction with high accuracy [5].

A stock market is an open market or public market for buying and selling shares for publicly listed companies. Different types of stocks are known for equities and ownership in the company. Mediator for buying and selling of shares monitored by the stock exchange [6]. The share market hugely impacts customers, businessmen, and financial institutions. In present days advanced technologies are used for data classification and analysis purposes [7]. Using technical models in financial institutions for price prediction, investment opportunities, investment portfolio optimization, etc [8].

Figure 1: Sample image for stock market

The following paper explains section 1 for introduction. Section 2 for the proposed architecture. Section 3 for results and analysis and comparative study. Section 4 paper conclusion.

II. PROPOSED ARCHITECTURE

A stock market is an open market or public market for buying and selling shares for publicly listed companies. Different types of stocks are known for equities and ownership in the company. Mediator for buying and selling of shares monitored by the stock exchange [9]. The share market hugely impacts customers, businessmen, and financial institutions. In present days advanced technologies are used for data classification and analysis purposes. Using technical models in financial institutions for price prediction, investment opportunities, investments, and portfolio optimization, etc [10].

A hybrid model means a combination of simple machine learning algorithms to try to predict better results compared to existing work. In this process sometimes get better results and sometimes predict the bad results also [11]. Our research focuses on predicting the results. Elastic Net Regression is a versatile tool in the data scientist's toolkit, offering a robust way to handle complex datasets with multicollinearity and high dimensionality. Its ability to perform feature selection and manage the bias-variance trade-off makes it a valuable algorithm for many machine-learning tasks [12].

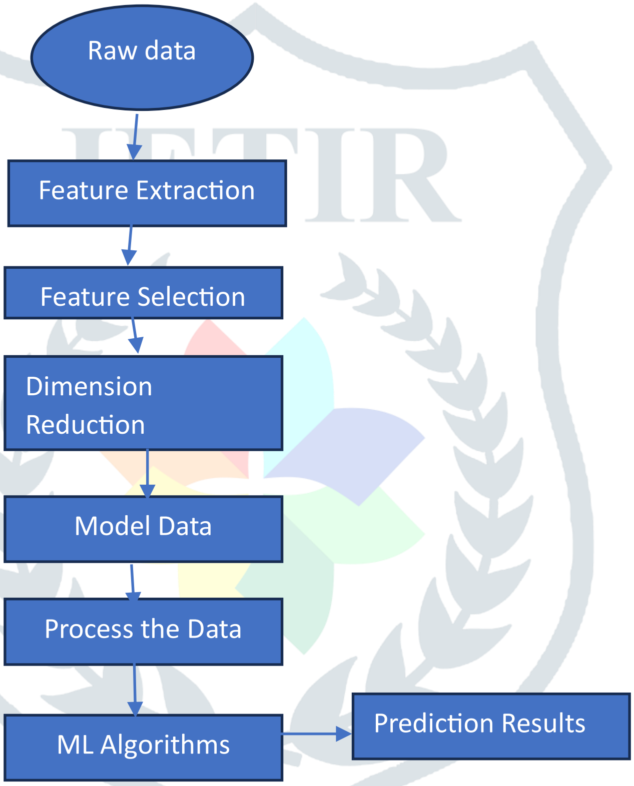

Figure 2: Architecture of the proposed solution [13]

2.1 Dataset description

The dataset has around 60 features which include features extracted from OHLC, other index prices such as Nasdaq-100 ETF& S&P 500, and technical Indicators such as Bollinger bands, and EMA, all of these features have something to offer for forecasting. Few expresses about the trend is a signal, that the stock is overbought, and few portray the strength of the price trend. In this paper, we will analyze the data and create dissimilar machine learning models like Linear Regression, Ridge Regression, Lasso Regression, and Elastic Net Regression model to forecast Stock Prices. Annaly, form an ensembled model from any machine learning algorithms mentioned above.

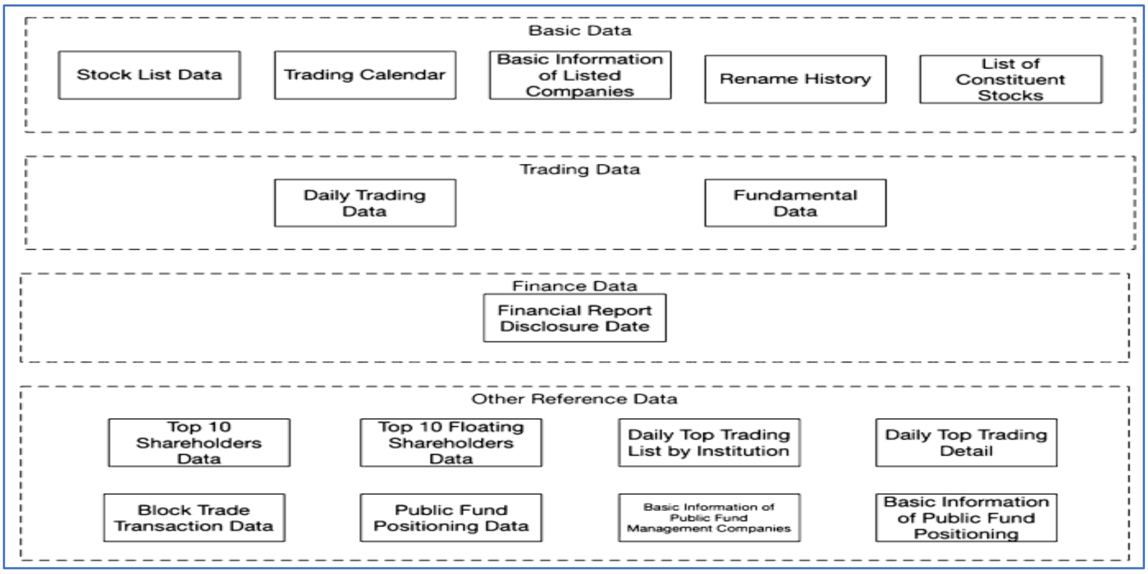

Figure 3: A data structure for the extracted dataset

Table 1: Dataset overview table with different categories and subsets of fields

Data table name	Category	Field
Stock list	Basic data	Stock ID, Stock name, Geographic info, Industry, Full name, English name, Market type, Stock exchange ID, Currency, List status, List date, Delist date, If the stock is HS constituent
Trading calendar	Basic data	Stock exchange ID, Calendar date, If the date is open for trading, Pervious trading date
Basic information of listed companies	Basic data	Stock ID, Stock exchange ID, Corporate representative, General manager, Secretary, Authorized capital, Registration date, Province, City, Introduction, Website, Email, Office address, Number of employees, Main business, Business scope
Renamed history	Basic data	Stock ID, Stock name, Start date, End date, Announcement date, Rename reason
Constituent stock information	Basic data	Stock ID, Constituent type, Included date, Excluded date, If the stock is new
Daily trading data	Trading data	Stock ID, Trading date, Opening price, Highest price, Lowest price, Closing price, Previous closing price, Price change, Price change percentage, Volume, Amount
Fundamental data	Trading data	Stock ID, Trading date, Closing price, Turnover rate, Free turnover rate, Volume ratio, Price-to-earning ratio, Price-to-earning ratio TTM, Price-to-book ratio, Price-to-sales ratio, Price-to-sales TTM, Total share capital, Circulating shares, Tradable circulating shares, Aggregate market value, Circulation market value
Financial report disclosure date	Finance data	Stock ID, Latest disclosure date, Reporting period, Scheduled disclosure date, Actual disclosure date, Disclosure modification date
Top 10 shareholders data	Other reference data	Stock ID, Announcement date, End date, Shareholder name, Holding amount, Holding ratio
Top 10 floating shareholders data	Other reference data	Stock ID, Announcement date, End date, Shareholder name, Holding amount
Daily top trading list by institution	Other reference data	Stock ID, Trading date, Institution name, Trading amount—buy, Trade ratio—buy, Trading amount—sell, Trade ratio—sell, Net turnover
Daily top transaction detail	Other reference data	Stock ID, Trading date, Stock name, Closing price, Price change percentage, Turnover rate, Amount—overall, On-list amount—sell, On-list amount—buy, On-list turnover, On-list net trading amount, On-list net trading ratio, On-list net turnover ratio, Circulation market value, Reason
Block trade transaction data	Other reference data	Stock ID, Trading date, Price, Volume, Amount, Buyer, Seller
Public fund positioning data	Other reference data	Fund ID, Announcement date, End date, Stock ID, Market value, Volume, Market value ratio, Circulation market value ratio

III. RESULTS AND ANALYSIS

3. 1 Importing required libraries

Our project use Jupiter notebook for python programming environment. The following libraries instal for our project implementation.

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from sklearn import metrics

from sklearn.linear_model import LinearRegression
from sklearn.linear_model import Ridge
from sklearn.linear_model import Lasso
from sklearn.linear_model import ElasticNet

from sklearn.ensemble import VotingRegressor


from sklearn.metrics import mean_squared_error

%matplotlib inline
import warnings
warnings.filterwarnings("ignore")
```

3.2 Load the data

The next step is loading the stock price dataset.

Table 2: Sample loading data

Open	High	Low	Close	Volume	SD20	Upper_Band	Lower_Band	S_Close(t-1)	S
6.66	6.69	6.50	6.60	154208600	0.169237	6.827473	6.150527	6.67	
6.57	6.66	6.44	6.45	152397000	0.168339	6.819677	6.146323	6.60	
6.43	6.78	6.32	6.78	252170800	0.180306	6.861112	6.139888	6.45	
6.72	6.97	6.71	6.93	339440500	0.202674	6.931847	6.121153	6.78	
7.02	7.03	6.83	6.87	199181500	0.216680	6.974860	6.108140	6.93	

Time

Figure 4: Plot Time Series chart for AAPL

3.3 Pre-Processing of Data

Remove a few columns, they are not required to our practical experiment [14].

```
df_Stock = df_Stock.drop(columns='Date_col')
```

3.4 Test &Train Set

Close_forecast is the column that we are trying to predict here which is the price for the next day.

```
X_train, X_val, X_test, Y_train, Y_val, Y_test = create_train_test_set(df_Stock)

    Historical Stock Data length is -  3732
    Training Set length -  3284
    Validation Set length -  373
    Test Set length -  74
    (3284, 61) (373, 61) (75, 61)
    (3284,) (373,) (75,)
```

3.5 Model building and comparison

3.5.1 basic utility functions

```
print(' ')

print("Test R-squared: ",round(metrics.r2_score(Y_test,Y_test_pred),2))
print("Test Explained Variation: ",round(metrics.explained_variance_score(Y_test,Y_test_pred),2))
print('Test MAPE:', round(get_mape(Y_test,Y_test_pred), 2))
print('Test Mean Squared Error:', round(metrics.mean_squared_error(Y_test,Y_test_pred), 2))
print("Test RMSE: ",round(np.sqrt(metrics.mean_squared_error(Y_test,Y_test_pred)),2))
print("Test MAE: ",round(metrics.mean_absolute_error(Y_test,Y_test_pred),2))
```

A. Linear Regression

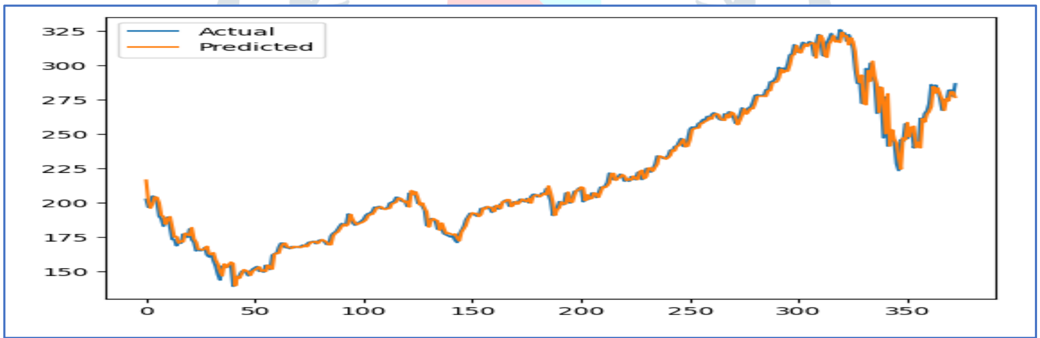

Figure 5: Linear regression Predictions

B. Ridge Regression

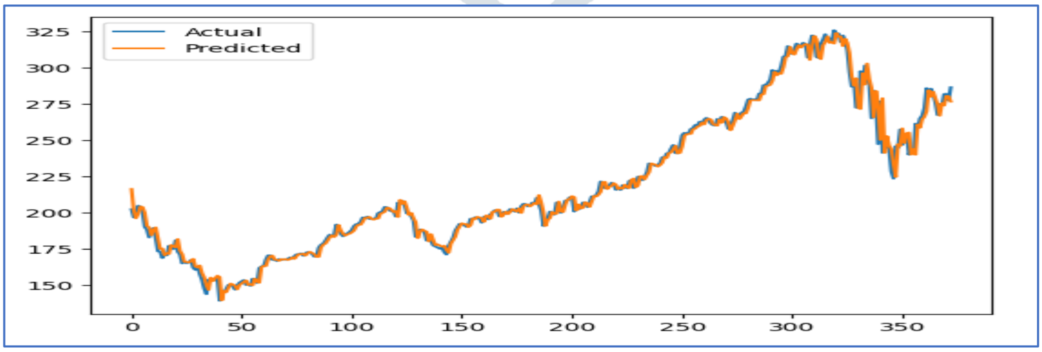

Figure 6: Ridge regression Predictions

In machine learning, ridge regression assistances decrease overfitting that consequences from model difficulty. Model complication can be owing to: Features are the model's analysts and parameters in machine learning [15].

C. Lasso Regression

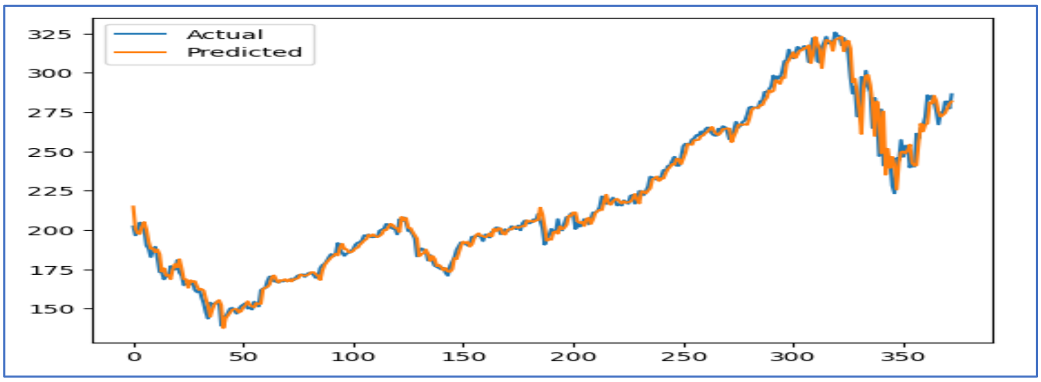

Figure 6: lasso regression Predictions

D. Elastic Net Regression

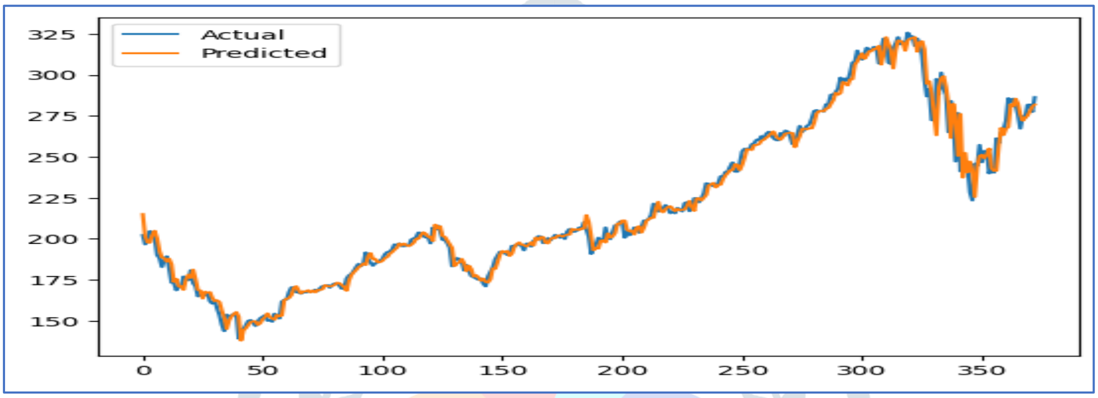

Figure 7: Elastic Net regression Predictions

3.6 Comparing various machine learning algorithms

MSE is used to check how close estimates to actual values. Lower the MSE, the closer is forecast too actual. This model assessment measure for regression models and the lesser value designates a better fit.

Table 3: MSE Predictions comparison

	Model	MSE
0	Linear Regression	79.213671
1	Ridge Regression	79.157075
2	Lasso Regression	89.400615
3	Elastic Net Regression	84.036811

The above comparison shows that Lasso regression gives more MSE, so it car far away from the actual. Linear regression and Ridge regression both algorithms show the same MSE, they predict as actual [16].

5.6.3 Building Ensemble Machine Learning Model

Individual mean square errors:

Model MSE

0 Linear Regression 79.213671

1 Ridge Regression 79.157075

2 Lasso Regression 89.400615

3 Elastic Net Regression 84.036811

A Mean Squared Error (MSE) value of 77.58 for an ensemble model combining Linear Regression, Ridge Regression, and Elastic Net Regression suggests that the model's predictions exhibit a moderate level of error when compared to the true target values. Finally, ensemble learning gives more actual prediction compared to individual algorithm implementation.

IV. CONCLUSION

A stock market is an open market or public market for buying and selling shares for publicly listed companies. Different types of stocks are known for equities and ownership in the company. Mediator for buying and selling of shares monitored by the stock exchange [17]. The share market hugely impacts customers, businessmen, and financial institutions. In present days advanced technologies are used for data classification and analysis purposes. Using technical models in financial institutions for price prediction, investment opportunities, investments, and portfolio optimization, etc [18]. Our present research uses machine learning models for prediction for financial trend analysis. It shows a comprehensive review of practical applications of machine learning models. A Mean Squared Error (MSE) value of 77.58 for an ensemble model combining Linear Regression, Ridge Regression, and Elastic Net Regression suggests that the model's predictions exhibit a moderate level of error when compared to the true target values.

REFERENCES

1. Atsalakis G S, Valavanis KP. Forecasting stock market short-term trends using a neuro-fuzzy-based methodology. Expert Syst Appl. 2009;36(7):10696–707.
2. Ayo CK. Stock price prediction using the ARIMA model. In: 2014 UK Sim-AMSS 16th international conference on computer modeling and simulation. 2014.
3. Brownlee J. Deep learning for time series forecasting: predict the future with MLPs, CNNs, and LSTMs in Python. Machine Learning Mastery. 2018.
4. Eapen J, Bein D, Verma A. Novel deep learning model with CNN and bi-directional LSTM for improved stock market index prediction. In: 2019 IEEE 9th annual Computing and Communication Workshop and Conference (CCWC). 2019. pp. 264–70.
5. Fischer T, Krauss C. Deep learning with long short-term memory networks for financial market predictions. Eur J Oper Res. 2018;270(2):654–69.
6. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Mach Learn 2002; 46:389–422.
7. Hafezi R, Shahrabi J, Hadavandi E. A bat-neural network multi-agent system (BNNMAS) for stock price prediction: a case study of DAX stock price. Appl Soft Comput J. 2015; 29:196–210.
8. Halko N, Martinsson PG, Tropp JA. Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions. SIAM Rev. 2001;53(2):217–88.
9. Hassan MR, Nath B. Stock market forecasting using Hidden Markov Model: a new approach. In: Proceedings—5th international conference on intelligent systems design and applications 2005, ISDA'05. 2005. pp. 192–6.
10. Hochreiter S, Schmid Huber J. Long short-term memory. J Neural Comput. 1997;9(8):1735–80.
11. Mr. Sujan Reddy, Ms. Renu Sri, and Subhani Shaik," Sentimental Analysis using Logistic Regression", International Journal of Engineering Research and Applications (IJERA), Vol.11, Series-2, July-2021.
12. Ms. Mamatha, Srinivasa Datta and Subhani Shaik," Fake Profile Identification using Machine Learning Algorithms", International Journal of Engineering Research and Applications (IJERA), Vol.11, Series-2, July-2021.
13. J. Lavanya, M. Ramesh, J. Sravan Kumar, G. Rajaramesh and Subhani Shaik," Hate Speech Detection Using Decision Tree Algorithm", Journal of Advances in Mathematics and Computer Science, Volume 38, Issue 8, Page 66-75, June-2023.
14. Neeraja, Anupam, Sriram, Subhani Shaik and V. Kakulapati," Fraud Detection of AD Clicks Using Machine Learning Techniques", Journal of Scientific Research and Reports, Volume 29, Issue 7, Page 84-89, June-2023.
15. P. Pranathi, V. Revathi, P. Varshitha, Subhani Shaik and Sunil Bhutada," Logistic Regression Based Cyber Harassment Identification", Journal of Advances in Mathematics and Computer Science, Volume 38, Issue 8, Page 76-85, June-2023.
16. Dr. K Vijayalaxmi, Dr. Subhani Shaik, B Pramod, B Ramulu, P Sankeerth," Vegetable Price Prediction against Temperature Changes Using Machine Learning Techniques", Dickensian journal, volume 22, issue 6, June, 2022.
17. Dr. Sunil Bhutada and Subhani Shaik, "IPL Match Prediction using Machine Learning", IJAST, Vol.29, Issue 5, April-2020.
18. Shiva Keertan J and Subhani Shaik," Machine Learning Algorithms for Oil Price Prediction", International Journal of Innovative Technology and Exploring Engineering, Volume-8 Issue-8, 2019.