AMAZON STOCK PRICE PREDICTION BY USING SMLT

S.M. Ajitha[1], S.Karthick[2], Mu.Karthikeyan[3], N.T.Naveen Balaji[4]

[1] *(Assistant Professor) Department of Information Technology, Meenakshi College of Engineering, Chennai, Tamil Nadu, India*

[2-4]*Students of Department of Information Technology, Meenakshi College of Engineering, Chennai, Tamil Nadu,*

---***---

Abstract - *Stock prices are determined by a company's initial public offering (IPO). Investment firms use a variety of metrics to determine what the stock's price should be. Traders use financial metrics to determine the value of the company, including its history of earnings, changes in the market, and the profit it can reasonably be expected to bring in. Stock price prediction has become an important research area. We propose a SMLT algorithms to accurately predict the stock price. Linear regression algorithm has the best accuracy with precision, Recall and F1 Score.*

Keywords: **LINEAR REGRESSION, ACCURACY, PRECISION**

1. INTRODUCTION:

Amazon (AMZN) is one of the most well-known companies in the world and holds a resolute position. The stock market prediction problem has drawn a lot of interest from both scholars and practitioners since it is crucial but difficult. An RNN-based ensemble model for financial market forecasting using news releases [1]-[2] was described in the existing solution. The most representative elements from financial news and historical data were culled using sentiment analysis and the sliding window approach. By using data from the past, machine learning can forecast the future [3]. Computers may learn without being explicitly taught thanks to a technique called machine learning (ML), which is a subset of artificial intelligence (AI). Algorithms with specialized functions are used during the training and prediction processes. The objective is to build a machine learning model for forecasting the price of Amazon stock [4] that might ultimately replace the supervised machine learning regression models, which can be updated by predicting outcomes with the highest level of accuracy by contrasting supervised methods [10]. The rest of this article is divided into the following sections. The three algorithms (KNN, EN, and SVR) and the structure of our suggested LR architecture are explained in Section II. In Section III, experiments and analyses are described. This includes data preparation, the evaluation index, experimental outcomes, and statistical analysis. In Section IV, conclusions are reached.

1.1 Existing System:

When making stock market investing selections, traders and investors can benefit greatly from financial news announcements. The stock market prediction problem has drawn a lot of interest from academics and industry professionals since it is crucial yet difficult. Due to the complexity and ambiguity of the natural language [1] [15][16] used in the news, conventional machine learning models frequently fail to understand the substance of financial news. An RNN-based ensemble [4] [5] [6] model for financial market prediction using news releases was reported in this study. The sliding window approach and sentiment analysis were used to extract the most representative characteristics from historical data and financial news. Compared to conventional pre-processing techniques (such bag-of-words and TF-IDF), which extract tens of thousands of features, this significantly decreased the number of dimensions.

1.2 Proposed System:

Data gathered from a variety of sources make up the proposed project. The data will be examined for accuracy. The cleansed data will be created for testing and training purposes. The machine learning approach is used to generate the model. To identify the optimal method, many algorithms are used. The ideal one serves as a template. The new stock price is predicted using the data model.

2. ALGORITHM USED:

The Elastic net, KNN, and SVR are three well-known algorithms capable of handling sequential structural data. The design of our LR Algorithm, which has the highest accuracy for stock price forecasts, is then described.

2.1: K-NEAREST NEIGHBORS:

A straightforward technique known as "K nearest neighbors" stores all of the relevant data and forecasts the numerical objective based on a similarity metric. (e.g., distance functions) [7]-[9]. Since the first decade of the 1970s, KNN has been utilized as a non-parametric method for statistical estimation and pattern identification. The supervised machine learning technique known as k-nearest neighbor's (KNN) is straightforward and simple to use. It may be used to tackle classification and regression issues. The KNN method pre supposes the presence of nearby comparable objects. Alternatively said, related objects are [9] close to one another. With some mathematics we may have learnt in our early years—calculating the distance between points on a graph—KNN encapsulates the notion of similarity (also termed distance, proximity, or closeness).

```
MEAN ABSOLUTE ERROR VALUE IS: 455.373813322144

MEAN SQUARED ERROR VALUE IS: 583119.4605785768

MEDIAN_ABSOLUTE_ERROR VALUE IS: 235.8055054

ACCURACY RESULT OF KNN IS: 22.941430612593507

R2_SCORE VALUE IS: 0.2228417842314866
```

Fig-1: Predicted accuracy using KNN

2.2: SUPPORT VECTOR REGRESSION:

Machine learning classification and regression issues are frequently solved with Support Vector Machines (SVM) [20]-[22]. The classification problem is a generalization of the regression problem, in which the model produces an output with continuous values as opposed to one from a limited set. In other terms, a regression model calculates the estimate of a multivariate continuous function. SVMs formulate binary classification issues as convex optimization issues and use those to solve the issues. The goal of the optimization issue is to accurately categorize as many training points as feasible while determining the largest margin separating the hyperplane. Support vectors in SVMs are used to represent this ideal hyperplane. The SVM is adaptable to regression issues due to its sparse solution and strong generalization. In order to generalize SVM to SVR, the function is surrounded by a -insensitive zone known as the supervised learning context.

```
MEAN ABSOLUTE ERROR VALUE IS: 438.4622916016202

MEAN SQUARED ERROR VALUE IS: 805454.844288342

MEDIAN_ABSOLUTE_ERROR VALUE IS: 91.97620729048856

ACCURACY RESULT OF SVR IS: 5.855671096174575

R2_SCORE VALUE IS: -0.1095730957919121
```

Fig-2: Predicted accuracy using SVR

2.3: ELASTIC NET:

A model that assumes a linear connection between the input variables and the target variable is known as a "linear regression" model. This connection may be conceptualized as a hyperplane connecting the input variables to the target variable in higher dimensions and as a line with a single input variable [16]-[19]. An optimization procedure is used to find the model's coefficients in order to reduce the total squared error between the predictions (yhat) and the anticipated goal values. (y).

Loss = sum i=0 to n (y_i – yhat_i)^2

The predicted coefficients of the model might grow significantly, which makes the model sensitive to inputs and potentially unstable. This is especially true for issues when there are fewer observations (samples) or more samples (n) than input variables or predictors (p) (so-called p >> n issues).

A common punishment is to devalue a model according to the sum of the squared coefficient values. The punishment for this is an L2. Although it prevents any coefficients from being dropped from the model, an L2 penalty reduces the size of all coefficients.

L2_penalty = j=0 to p beta_j2

Another common punishment is to devalue a model according to the total of the absolute coefficient values. It is known as the L1 penalty. All coefficient sizes are reduced by an L1 penalty, which also permits certain coefficients to be minimized to zero, which eliminates the predictor from the **model.Sum j=0 to p abs l1_penalty(beta_j)**A penalized linear regression model called an elastic net incorporates both L1 and L2 penalties during training. The weight assigned to the L1 and L2 penalties is specified by the hyper parameter "alpha," which is taken from "The Elements of Statistical Learning." The L1 penalty's contribution is weighted using an alpha value between 0 and 1, and the L2 penalty's contribution is weighted using one minus the alpha value. **((1 - alpha) * l2_penalty) + (alpha * l1_penalty) = elastic_net_penalty**. For instance, an alpha of 0.5 would allow each penalty to contribute equally to the loss function. All weight is given to the L2 penalty with an alpha value of 0, while a value of 1 yields all weight to the L1penalty. Elastic net has the advantage of allowing a balance between the two penalties, which can lead to better performance on particular problems than a model with only one penalty. Lambda, a further hyperparameter, affects how much the total of both penalties is weighted in the loss function. The fully weighted penalty is applied by default with a value of 1.0; if the value is zero, the penalty is not applied. Lambda values as tiny as 1e-3 or less are typical.

Loss+(lambda*elastic_net_penalty)=elastic_net_loss.

```
MEAN ABSOLUTE ERROR VALUE IS: 5.7832568845229355

MEAN SQUARED ERROR VALUE IS: 203.60321749003285

MEDIAN_ABSOLUTE_ERROR VALUE IS: 1.3577014424583922

ACCURACY RESULT OF ELASTICNET IS: 99.97318555333567

R2_SCORE VALUE IS: 0.9997318009491324
```

Fig-3: Predicted accuracy using Elastic Net

2.4: LINEAR REGRESSION:

A machine learning algorithm built on supervised learning is linear regression. The job of predicting the value of a dependent variable (y) based on an independent variable is carried out using linear regression. (x). Thus, x (the input) and y (the output) are shown to be linearly related using this regression approach.(output). It completes a regression job. Regression creates a goal prediction value based on independent variables. It is mostly used to determine how variables and forecasting relate to one another. The type of link that different regression models take into account between the dependent and independent variables, as well as the amount of independent variables utilized, vary. The job of predicting the value of a dependent variable (y) based on an independent variable is carried out using linear regression. (x). Consequently, this regression approach identifies a linear connection between x(input) and y(output).

The linear regression procedure, often known as linear regression, illustrates a linear connection between one or more independent variables (y) and a dependent variable (y). As a result of displaying a linear connection, linear regression may be used to determine how the value of the dependent variable [23]-[25] changes in proportion to the value of the independent variable. A sloping straight line illustrating the relationship between the variables is provided by the linear regression model.

The general form of each type of regression is:

- **Simple linear regression:** $Y = a + bX + u$
- **Multiple linear regression:** $Y = a + b_1X_1 + b_2X_2 + b_3X_3 + ... + b_tX_t + u$

Where:

- Y = the variable that you are trying to predict (dependent variable).
- X = the variable that you are using to predict Y (independent variable).
- a = the intercept.
- b = the slope.
- u = the regression residual.

```
MEAN ABSOLUTE ERROR VALUE IS: 2.757359381236852

MEAN SQUARED ERROR VALUE IS: 46.0574345755943

MEDIAN_ABSOLUTE_ERROR VALUE IS: 0.6966165052120346

ACCURACY RESULT OF LR IS: 99.99371054136492

R2_SCORE VALUE IS: 0.9999370740424633
```

Fig-4: Predicted accuracy using Linear Regression

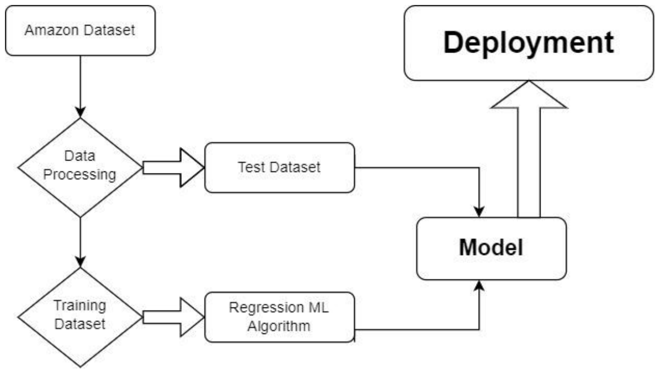

Fig-5: Architecture of Proposed system

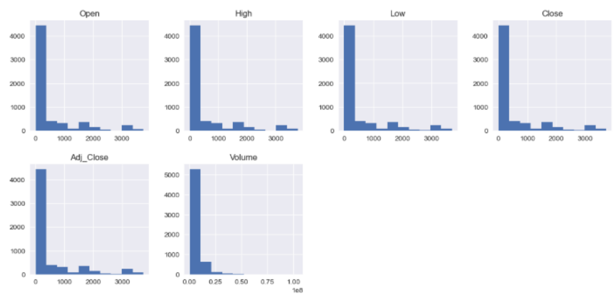

Fig-6: Histogram comparison of different values

2.5: DEPLOYEMENT:

DJANGO:

A Python-based server-side web framework with a wide range of features, Django, is very well-liked. Django is one of the most widely used web server frameworks, how to set up a development environment, and how to start using it to build your own web applications Regardless of the popular operating system you use, this article should provide you with all you need to begin building Django apps. A high-level Python web framework called Django enables the quick creation of safe and dependable websites. Django, which was created by seasoned programmers, handles a lot of the pain associated with web development, allowing you to concentrate on developing your app without having to recreate the wheel.

HTML:

Hyper Text Markup Language is referred to as HTML. Using a markup language, it is used to create web pages. Markup language and hypertext are combined to form HTML. The link between web pages is defined by hypertext. To specify the text document within the tag that specifies the structure of web pages, a markup language is employed. This language is used to annotate (create computer-readable comments on) material so that a computer can comprehend it and change the content as necessary. Most markup languages, including HTML, are understandable by humans. Tags are used in the language to specify what text processing is required.

CSS:

Cascading Style Sheets are known as CSS. It is a language for defining how Web pages are presented, including the colors, layout, and fonts that make our web sites appealing to users. For creating style sheets for the web, CSS was created. Any XML-based markup language may be used with it because it is not dependent on HTML. Let's attempt to decode the acronym now:

• Cascading: Falling into Old Habits

• Sheets: Writing in various documents using our style

3. CONCLUSION:

Customers may buy a range of products and services from Amazon.com, Inc. Through its shops, customers may purchase goods from both its own third-party sellers and those it has bought to resell. Additionally, it manufactures and sells electrical equipment. The project's primary objectives are to determine accuracy, reduce mistake rates, and obtain results from deployment. Data cleansing and processing, missing value analysis, exploratory analysis, and model creation and assessment came next in the analytical process. The best accuracy on the public test set with the highest accuracy score will be discovered. By employing the algorithm with the highest degree of accuracy, this can assist in determining the Amazon Stock Price.

4. FUTURE WORK:

To link the AI Model, forecast the price of Amazon.com. This procedure may be made automated by displaying the prediction results in a desktop or online application. To more efficiently do the job that has to be done in an AI environment.

REFERENCES

[1] Raymond Chiong , Zongwen Fan, Zhongyi Hu, and Sandeep Dhakal “A Novel Ensemble Learning Approach for Stock Market Prediction Based on Sentiment Analysis and the Sliding Window Method”,2022.

[2] T. Fischer and C. Krauss, “Deep learning with long short-term memory 600 networks for financial market predictions,” Eur. J. Oper. Res., vol. 270, 601 no. 2, pp. 654–669, Dec. 2017. 602

[3] X. Wu, Q. Ye, H. Hong, and Y. Li, “Stock selection model based on 603 machine learning with wisdom of experts and crowds,” IEEE Intell. 604 Syst., vol. 35, no. 2, pp. 54–64, Mar. 2020. 605

[4] M. Jiang, L. Jia, Z. Chen, and W. Chen, “The two-stage machine 606 learning ensemble models for stock price prediction by combining 607 mode decomposition, extreme learning machine and improved har- 608 mony search algorithm,” Ann. Oper. Res., vol. 309, pp. 553–585, 609 Jun. 2020. 610

[5] D. C. Furlaneto, L. S. Oliveira, D. Menotti, and G. D. C. Cavalcanti, 611 “Bias effect on predicting market trends with EMD,” Expert Syst. Appl., 612 vol. 82, pp. 19–26, Oct. 2017. 613

[6] I. E. Fisher, M. R. Garnsey, and M. E. Hughes, “Natural language 614 processing in accounting, auditing and finance: A synthesis of the 615 literature with a roadmap for future research,” Intell. Syst. Accounting, 616 Finance Manage, vol. 23, no. 3, pp. 157–214, Jul. 2016. 617

[7] G. G. Creamer, “Network structure and market risk in the Euro- 618 pean equity market,” IEEE Syst. J., vol. 12, no. 2, pp. 1090–1098, 619 Jun. 2018. 620

[8] S. Feuerriegel and J. Gordon, “Long-term stock index forecasting based 621 on text mining of regulatory disclosures,” Decis. Support Syst., vol. 112, 622 pp. 88–97, Aug. 2018. 623

[9] T. Geva and J. Zahavi, “Empirical evaluation of an automated intraday 624 stock recommendation system incorporating both market data and textual 625 news,” Decis. Support Syst., vol. 57, pp. 212–223, Jan. 2015. 626

[10] F. Z. Xing, E. Cambria, and R. E. Welsch, “Natural language based 627 financial forecasting: A survey,” Artif. Intell. Rev., vol. 50, no. 1, 628 pp. 49–73, Jun. 2018. 629

[11] K. Yuan, G. Liu, J. Wu, and H. Xiong, "Dancing with Trump in the 630 stock market: A deep information echoing model," ACM Trans. Intell. 631 Syst. Technol., vol. 11, no. 5, pp. 1–22, Oct. 2020. 632

[12] K. Nam and N. Seong, "Financial news-based stock movement prediction using causality analysis of influence in the Korean stock market," 634 Decis. Support Syst., vol. 117, pp. 100–112, Feb. 2019. 635 [12] J. Nikkinen and J. Peltomäki, "Crash fears and stock market effects: 636 Evidence from web searches and printed news articles," J. Behav. 637 Finance, vol. 21, no. 2, pp. 117–127, 2019. 638

[13] S. Evert et al., "Combining machine learning and semantic features in 639 the classification of corporate disclosures," J. Log., Lang. Inf., vol. 28, 640 no. 2, pp. 309–330, 2019. 6M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[14] E. Saquete, D. Tomás, P. Moreda, P. Martínez-Barco, and M. Palomar, 643 "Fighting post-truth using natural language processing: A review 644 and open challenges," Expert Syst. Appl., vol. 141, Mar. 2020, 645 Art. no. 112943. 646

[15] N. Seong and K. Nam, "Predicting stock movements based on financial 647 news with segmentation," Expert Syst. Appl., vol. 164, Feb. 2021, 648 Art. no. 113988. 649

[16] R. Chiong, G. S. Budhi, S. Dhakal, and F. Chiong, "A textual-based 650 featuring approach for depression detection using machine learning 651 classifiers and social media texts," Comput. Biol. Med., vol. 135, 652 Aug. 2021, Art. no. 104499. 653

[17] S. F. A. Zaidi, F. M. Awan, M. Lee, H. Woo, and C.-G. Lee, 654 "Applying convolutional neural networks with different word repre655 sentation techniques to recommend bug fixers," IEEE Access, vol. 8, 656 pp. 213729–213747, 2020. 657

[18] S. M. A. Islam, B. J. Heil, C. M. Kearney, and E. J. Baker, "Protein 658 classification using modified n-grams and skip-grams," Bioinformatics, 659 vol. 34, no. 9, pp. 1481–1487, May 2018. 660

[19] S. Yilmaz and S. Toklu, "A deep learning analysis on question classi661 fication task using Word2vec representations," Neural Comput. Appl., 662 vol. 32, pp. 2909–2928, Jan. 2020. 663

[20] J. Lilleberg, Y. Zhu, and Y. Zhang, "Support vector machines and 664 Word2vec for text classification with semantic features," in Proc. IEEE 665 14th Int. Conf. Cognit. Informat. Cognit. Comput. (ICCI*CC), Jul. 2015, 666 pp. 136–140. 667

[21] Y. Qiu, H.-Y. Yang, S. Lu, and W. Chen, "A novel hybrid model based 668 on recurrent neural networks for stock market timing," Soft Comput., 669 vol. 24, no. 20, pp. 15273–15290, 2020. 670

[22] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in 671 deep learning based natural language processing," IEEE Comput. Intell. 672 Mag., vol. 13, no. 3, pp. 55–75, Aug. 2018. 673

[23] M. Kraus and S. Feuerriegel, "Decision support from financial disclo674 sures with deep neural networks and transfer learning," Decis. Support 675 Syst., vol. 104, pp. 38–48, Dec. 2017. 676

[24] A. Atkins, M. Niranjan, and E. Gerding, "Financial news predicts stock 677 market volatility better than close price," J. Finance Data Sci., vol. 4, 678 no. 2, pp. 120–137, Jun. 2018. 679

[25] Y. Liu, Y. Chen, S. Wu, G. Peng, and B. Lv, "Composite leading search 680 index: A preprocessing method of internet search data for stock trends 681 prediction," Ann. Oper. Res., vol. 234, no. 1, pp. 77–94, Nov. 2015.