Laptop Price Prediction using Machine Learning Algorithms

Mohammed Ali Shaik
Department of Computer Science &Artificial Intelligence,
SR University, Warangal, Telangana State, India
niharali@gmail.com

Nagamalla Sanjana
Department of Computer Science & Engineering,
SR University, Warangal, Telangana State, India
19k41a05d8@sru.edu.in

Medicherla Varshith
Department of Computer Science & Engineering,
SR University, Warangal, Telangana State, India
19k41a05d5@sru.edu.in

Rama Sujith
Department of Computer Science & Engineering,
SR University, Warangal, Telangana State, India
19k41a05e0@sru.edu.in

Sanka SriVyshnavi
Department of Computer Science & Engineering,
SR University, Warangal, Telangana State, India
19k41a05g8@sru.edu.in

Abstract— **The laptop has grown to be one of the most essential and used gadgets in our day-to-day existence for different activities. We will be supplied with many specs and company names in the market, it will become difficult for laptop computer makers to sell their merchandise and for customers to pick out one. Machine learning (ML) is high quality in assisting in making decisions and predictions from the large volume of facts produced. We have additionally viewed ML strategies being used in recent developments in the Internet of Things (IoT) areas. Various studies supply solely a glimpse into predicting the price of the laptop with ML techniques as in this paper, we suggest a novel technique that targets identification process through tremendous elements using making use of desktop getting to know fashions resulting in improving the accuracy in the prediction of laptop price. The prediction model is delivered with one-of-a-kind combos of features and several regarded computing device learning models. We are the use of a one-of-a-kind laptop to gain knowledge of fashions like Decision trees, Multiple linear regression, KNN, and Random forest to test which desktop mastering model is more accurate in predicting the rate of the laptop.**

Keywords— Machine learning, Data Mining, prediction model, Classification, Multiple linear regression

I. Introduction

The importance of laptops in our each day lives cannot be overstated. Humans evaluate and undertake science daily in everyday life. Laptops, like computer computers, are ubiquitous in our everyday lives. Raw statistics are changed into valuable and meaningful information in all of these areas: companies, banks and universities, and homes. Engineers, students, teachers, commercial enterprise people, and authorities businesses all use it for particular duties, amusement, web earnings, and office work. Laptops are a blessing for people if they are used for proper causes. The majority of companies maintain all of their operational and accounting statistics on a laptop. Today, a diploma in international enterprise operations except for the use of pc applied sciences and the web is unthinkable. Laptops make it simple to save and continue data, as well as come across archives and essential information. Excel sheets, documents and administrative systems, transaction processing, and traceability systems are just a few examples of equipment

and software that make administrative work easier and greater productive [21]. The company's laptops can compute and summarise statistics entered in reports, statements, and papers that are utilized for a range of purposes. Databases are maintained on laptops by organizations of all sizes. Databases are pc programs that store and manipulate essential data, such as patron contact information.

The "Machine Learning" is a department of "Artificial intelligence" that offers implementation purposes that can make future predictions based on past data. If you are a facts science enthusiast or a practitioner then this article will assist construct your own end-to-end desktop studying project from scratch. There are a variety of steps involved in building a laptop studying undertaking but not all the steps are obligatory to use in a single project, and it all depends on the data. In this article, we will construct a Laptop charge prediction challenge and examine the laptop by studying its mission lifecycle.

In machine learning, a multiple regression model is a specific kind of "supervised learning technique" for estimating the mapping feature to reap excellent predictions. Regression's primary motive is to construct an efficient model that can predict the based attribute from a list of attribute variables.

II. Literature Review

Paper – [1]: In "Classification Based on Decision Tree Algorithm for Machine Learning" Behzad Tahaz Jijo1*, Adnan Mohsina Abdulazeez proposed the basic goal of the Decision tree algorithm which is related to the supervised getting to know algorithm family, is building a coaching model that can be used to predict the class or cost of goal variables by using inferring decision policies from coaching data. Regression and classification problems can each be solved with the use of the "Decision Tree" [2] technique. The Decision Tree is ordinarily used for grouping purposes. Furthermore, the Decision Tree is a common labelling or classifying model in the field of "Data Mining" [3]. Where in each of the tree comprises of distinct set of nodes and branches where each of the subset tends to define the price of each of the node which represents each node to be the facet in a specific category which is to be categorized. Decision

trees offer a wide variety of functions due to their easy analysis and precision across many statistics types.

Paper–[2]: In "Decision tree methods: applications for classification and prediction" paper Ying Stanford University

| SU Department of Data Science Ph.D. Yan-Yan Song's Shanghai Jiao Tong University proposed that: A frequent facts mining approach for evolving various classification structures that are clearly based on distinct covariates for evolving the prediction algorithms which is the goal variable comprising of the tree methodology where the population set is classified into distinct branch like segments for shaping the inverted tree comprising of "a root node", "interior nodes", and "leaf nodes" [18]. The implemented algorithm is a purely non parametric due to which it is capable of generating a huge and complicated datasets that are required for imposing a complicated parametric framework. The major study is implemented on acquiring the information which has to be bifurcated into "training" and "validation" datasets as soon as the sample dimension is massive enough [22]. While constructing the "decision tree model" implemented over a coaching dataset for identifying the possible right tree dimension using the validation dataset for acquiring the final selected model [23]. One of the most essential and powerful algorithms in prediction is the Decision Tree. It's simple to anticipate the result for future data the usage of the tree model got from prior data [24].

Paper – [3]: In the paper "Popular Decision Tree Algorithm of Data Mining Techniques" Radwan Abbas F. H. Alhaurin, Ali S. A. Al-Haboob were discussed the selection tree is performed based on the quantity of data by imparting the effectiveness over the methodology adopted over the amount of memory that is accessible, as the Decision Tree classification algorithm can be thoroughly implemented using either in serial phases or in parallel phases. One of the logical model is the serial tree approach which is used to construct the modal by the use of a coaching information set and is represented as a binary tree. Tools and techniques: Jupyter notebook, Decision Tree.

Paper – [4]: Random Forest is made up of various special tree shape nodes. There are many interesting things in Random Forest. Random Forest has seen vast utilization in categorization, prediction, and "regression". Where the "Random Forest" algorithm affords a lot of blessings over common algorithms where "Random Forest" has a vast variety of applications. Traditional machine studying techniques generally result in low classifier accuracy and are susceptible to overfitting [14]. Many human beings study algorithms for merging classifiers to enhance accuracy. Many researchers start their research to improve the accuracy of classification by combining various classifiers. As LeoBreiman, has proposed an advanced Bagging algorithm which is in the early stages of the algorithm implementation [15].

Paper – [5]: the University of California, Berkeley, CA 94720 published a paper "Growing a group of trees and allow them to vote for the most famous type" which resulted in noteworthy increases while performing classification accuracy as the random vectors will tend to regulate each of the enhanced tree while ensembleing the frequently built ensembles that have grown. Bagging is an early example, in which a random choice without performing replacement of the selected instances while educating the set of each of the tree development process [12, 28].

They outline a huge wide variety of geometric facets and search thru a "random decision" of them for the fantastic break up at every node in an important paper on written personality recognition [13,29]. This remaining paper has affected my thinking. In all of these techniques, a random vector okay is created for the kth tree, unbiased of the previous "random vectors" denoted as 1,...,k1, however with the equal distribution as the tree has to be shaped usage the coaching set and k will result into a classifier through h(x, k) wherein x is an enter vector that is considered. For example, in bagging operation performed over random vector generated through counting of variety of darts that are thrown at randomly N boxes where the place N comprises of distinct examples over a training set [19]. Several independent random integers between (1, K) are merely used for performing random split in the selection of various dimensions and characterisation process determined through a software while performing tree construction process created at a massive rate using various strategies regarded to be random forests [20].

Paper – [6]: Celine Vans Department of Computer Science, Kotolike University Belgium said that A "random forest" area is a group of random "Decision Tree" nodes that combine the predictions of individual nodes to make predictions. Different methods for introducing randomization into the selection tree construction manner exist. A random forest can be used to forecast goal residences that are nominal (Classification) or numeric (Regression). One of the most wonderful forecasting models is random forests [25].

Paper – [7]: "A Brief Review of Nearest Neighbor Algorithm for Learning and Classification "Khalid Alkhatib, and Hassan Najadat say that the KNN is a machine learning approach that is very easy to use. A similarity-based categorization can be used to map the prediction problem [11]. A set of vectors is created from the historical and test data. Each feature is represented by an N-dimensional vector. Then, to make a choice, the same measure such as Euclidean distance is computed [16]. KNN is a type of lazy learning in which no mannequin or feature is constructed previously and as a substitute returns the k information from the education information set that is the same as the test [17].

Paper – [8]:, Suraj, and Surendra Babu K N proposed stock prediction the usage of the Knn algorithm is the "k-nearest neighbor" algorithm is a supervised algorithm that is easy to construct. KNN is also characterized as the laziest studying method, as it no longer strengthens a mannequin or functionality before returning the closest ok values from the training dataset to take a look at the data. The category label is then allotted to the take a look at statistics when a majority vote is received amongst the chosen okay values [26].

Paper – [9]: Kashvi Trunk, Sanjukta De, Srishti Verma, and Aleena were said that When precise approximations are implemented there is a chance that comprises of unknown or complex to estimate KNN classification was once created to operate characteristic analysis as the non parametric method performs categorization of patterns using the well recognized KNN Rule [27].

Paper – [10]: Naalla Vineeth and Maturi Ayyappa proposed MLR is an upgrade of "simple linear regression" which permits two or more "predictor variables" for

generating models response through a variable y which is a linear function, whereas easy linear regression only has one. The most frequent kind of linear regression evaluation is more than one linear regression. Much linear regression is a predictive evaluation that indicates the hyperlink between a continuous established variable and more than one independent factor [30].

III. PROBLEM DEFINATION

The problem assertion is that if a consumer needs to buy a laptop, our utility ought to be able to deliver an estimated price for a laptop primarily based on the user's preferences.

A. DATASET

For this project, we have obtained my dataset from Kaggle which contains 1302 rows of data with 12 columns denoting features and focus on developing the proposed prediction model using the 12 attributes to predict the price of the laptop.

B. DATA PRE-PROCESSING

Initially processing the data is an important task for a project in cleansing the information and making it appropriate for a machine-learning model which also increases the accuracy rate and efficiency of a laptop model. In this particular section, we relabel & convert some categorical features into numeric values. This is crucial for training machine learning models since machine learning models accept numeric values. We have a total of 1302 rows and 12 columns (attributes) in the dataset We also do feature engineering of CPU.
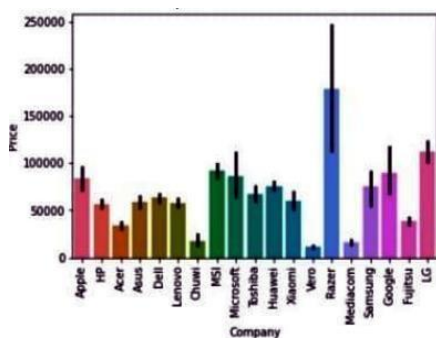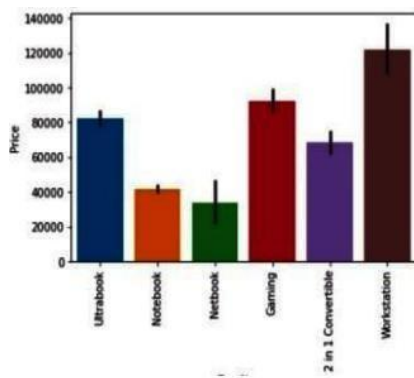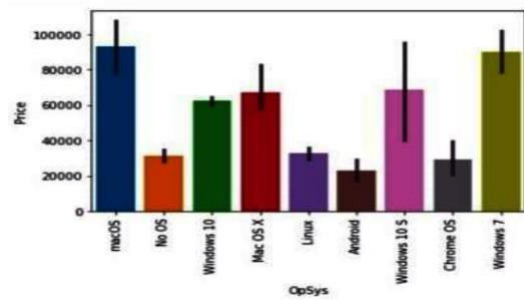
Fig.1. company attribute
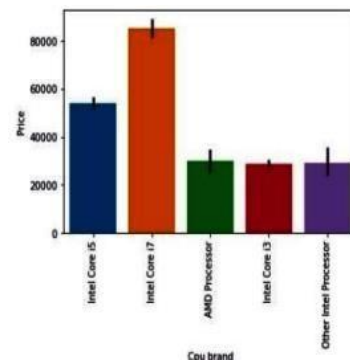
Fig.2. Type of Laptop

Fig.3. Operating System

Fig.4. Processor

From Figure 1 to Figure 4 are the results generated after the performing pre processing of attributes over the dataset which comprises of attributes considered based don figure 1 for company, figure 2 for type of laptop conserved, figure 3 for selected Operating System and finally figure 4 denotes the Processor type selected .

C. ALGORITHMS

We used many "Machine Learning (ML)" models comprising of distinct "Regression models" for solving our Prediction problem by making the data ready for performing training and testing of our ML model.

- Decision Tree: This algorithm is the phase of supervised gaining knowledge of the list of algorithms, and it is most important thing is to build a coaching mannequin that can be used for predicting the classification or fee of target variables via mastering user policies inferred from the coaching data. The Decision tree algorithm can be used to clear up regression and classification issues. Moreover "Decision Tree" is used for implementing the classification model using "Data Mining (DM)" approach where nodes and branches comprise of tree model. Where each of the nodes represent elements in a class are to be classified over a specific subset that defines the price through nodes as they are easy to evaluate and their precision on a couple of information forms, selection bushes have determined many implementation fields. Decision tree nodes are acknowledged for the superior view of performance outcomes. Because of their sturdy precision, optimizing split parameters, and greater tree pruning strategies like(ID3, and C4.5) are normally used through all identified information classifiers. The separate datasets are used in the training samples

228

from large datasets, which influences the precision of the test set.

- Random Forest: is a Group of various Decision trees that utilizes a "Laptops database" for implementing the import over initially trained version of the proposed network used for implementing training over thousands of Laptops data, as a result, will developed the library of most improvised features for representing accuracy of 87% and r2 score is 0.15% which are best compared to other algorithms.

- KNN: The prediction of the charge is computed by the usage of KNN as follows:

a) Obtain the range of nearer neighbours, k.

b) Evaluate the distance acquired between various neighbours and the question record.

c) Filter all coaching archives by the values of distance.

d) Using the majority of the category labels of k nearest neighbours, and assigning it as an estimation price of the question record.

IV. BUILDING THE MODEL

In this specific section, we relabel & convert some specific elements into numeric values. This is vital for education desktop mastering models when you consider that laptop studying fashions receive numeric values. We have a total of 1302 rows and 12 columns (attributes) in the dataset We additionally do characteristic engineering of CPU. So right here rather than display resolution, PPI (pixels per inch) has been taken. We cut upscreen decisions into x-resolution and y-resolution. From x-resolution and y-resolution we bought PPI(pixels per inch). After that screen decision is dropped and whilst splitting screen resolution we will get two columns for x-resolution and y-resolution. We will drop these two columns as well. Instead, we will have PPI(pixels per inch).

To evade any problems and error prone predictions, useless elements such as "Unamed:0", "Company" and "Product" will be eliminated from the dataset.

We would now extract and reorganize our statistics to better recognize the underlying factors that contribute to the fee of laptops.

If we take a appear at the display screen decision column, there appear to be laptops with touchscreen capabilities. Since touchscreen laptops are regarded to be greater high-priced than those beside them, a TouchScreen characteristic would be introduced to mark laptops with such capabilities.

We would then extract and substitute the display decision column with their respective pixel count.

Now comes the most tiring part of characteristic engineering, dealing with memory features. Upon nearer inspection, the memory column incorporates several kinds of reminiscence (SSD, HDD, SSHD, and Flash Storage). We would need to create 4 extra columns representing special memory sorts and extract their memory capacities individually. (Additional processing desires to be achieved for laptops having double memory configuration that makes use of the equal reminiscence types.(EX: 256GB SSD + 512GB SSD).

Having those memory configurations handled, we've decided to drop the GPU column completely as it includes a high variability of GPUs. Intel GPUs are built-in GPUs, and Nvidia GPUs are discrete whilst AMD GPUs are both integrated or discrete.

A. COMPARATIVE STUDY

We have collected the accuracies of the existing publications

TABLE I. TABLE TYPE STYLES

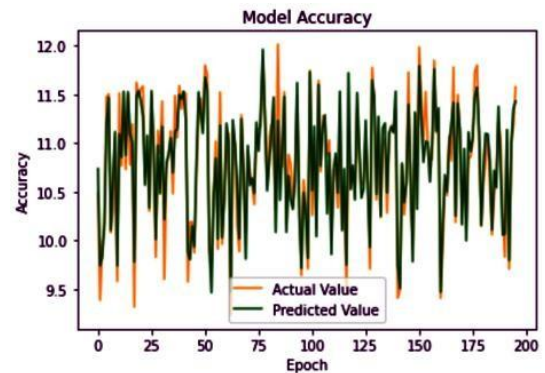| Sl.No.	Reference	Algorithm	Accuracy in Percentage
1	[1]	Decision tree	83.67
2	[1]	Random forest	88.25
3	[2]	KNN	79.93
4	[2]	MLR	79.52
5	[3]	SVM	80.83
6	[3]	Extra Trees	87.5
7	[4]	AdaBoost	79.29
8	[5]	RidgeRegression	81.2
9	[6]	Lasso Regression	80.7
10	[7]	XG BOOST	88.1
11	[8]	Stacking	88.1

B. PREDICTIONS

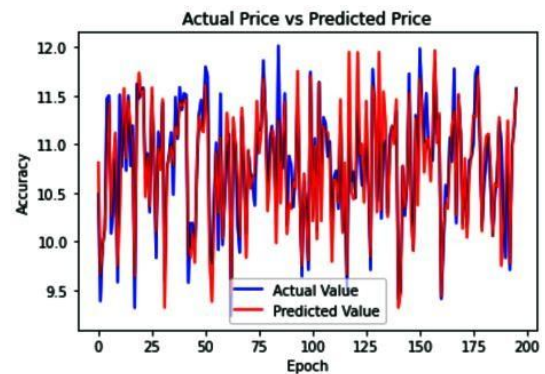Fig.5. Multiple linear regression Accuracy

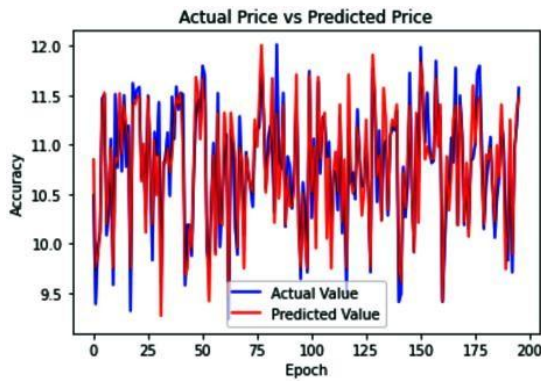

Fig.6. KNN Algorithm Accuracy
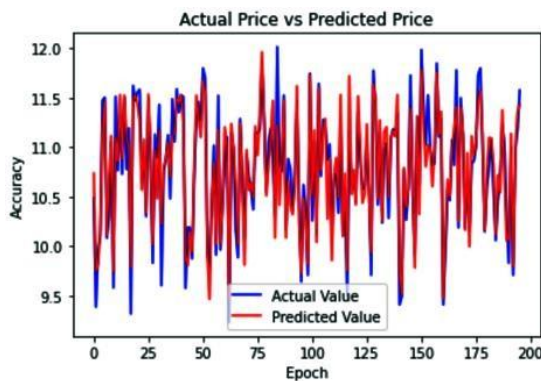
Fig.7. Decision Tree Accuracy

Fig.8. Random Forest Accuracy

The above figures represent the "Epoch vs Accuracy graph" which refers to the Model Accuracy of all the four different models that were implemented. We predict from test data; the following were the predictions made from four different models that were implemented.

Algorithm	Accuracy rate
Multiple Linear regression	80.73%
Random forest	88.75%
Decision tree	84.69%
KNN	80.21%

Price Prediction using	Accuracy	Mean absolute error
Linear Regression	0.8073277448418645	0.21017827976428802
Random forest	0.8875301429227068	0.15875464162243105
Decision tree	0.8469387025658242	0.18222142011995746
KNN	0.8021984604448553	0.19319716721521116

Fig.9 Accuracy and Mean absolute error rates

Fig.10. final result generated screen

The above figures represent the results obtained through the Laptop price predictor site The algorithms that we used linear regression, random forest, and decision tree, KNN. As we can see from the above table, we compared the accuracy rates of four different models. When we compared above these algorithms the accuracy rate for the Random forest is more than the other two algorithms. The above table shows the accuracy rates of the algorithms and also the mean absolute errors of the algorithms, whereas random forest is 88.65%, linear regression is 80.65%, KNN is 80.45% and the decision tree is 84.05%.

V.   CONCLUSION

We created a website that Predicts the prices of laptops using the user's desired configurations and using laptop price predictors and gaining knowledge of using the Decision Tree algorithm makes it easy for students, in particular in deciding the choice of laptop computer specifications that are most ideal for students to meet pupil desires and by the buying energy of students. Students no longer want to appear for various sources to discover laptop specs that are needed by college students in assembly the wishes of students because the laptop computer specifications from the results of the computer gaining knowledge of application have furnished the most perfect specs with the prices of their laptops.

REFERENCES

[1] Aminah Md Yusof and Syuhaida Ismail, Multiple Regressions in Analyzing House Price Variations. IBIMA Publishing Communications of the IBIMA Vol. 2012 (2012), Article ID 383101, 9 pages DOI: 10.5171/2012.383101.

[2] Babyak, M. A. What you see may not be what you get: A brief, nontechnical introduction to overfitting regression-type models. Psychosomatic Medicine, 66(3), 411-421.

[3] Atharva shoogles, Priyanka chair, Akshata gaud, Jinal Jain.House Price Forecasting using Data Mining Techniques International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Vol. 6, Issue 12, December 2017

[4] Darshan Sanghani, Kelby Erickson, and Mohammad Al Hasan Predicting Zillow Estimation Error Using Linear Regression and Gradient Boosting, IEEE 14th International Conference on MobilAd Hoc and Sensor Systems (MASS), Page(s):530 - 534.

[5] Model, Azmi Bin Khamis, Nur Khalidah Khalilah Binti Kamarudin, Comparative Study On Estimate House Price Using Statistical And Neural Network, International journal of scientific and technology, research volume 3, ISSUE 12, December 2014,Page(s):126-131.

[6] Adyan Nur Alfiyatin, Hilman Taufiq, Ruth Ema Febrita, Wayan Firdaus Mahmudy,Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization Case Study: Malang, East Java, Indonesia. (ITASCA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 10, 2017, Page(s):323-326..

[7] S. Jyothirmayee, V. D. Kumar, C. S. Rao, and R. S. Shankar, "Predicting Stock Exchange using Supervised Learning Algorithms," no. 1, pp. 4081-4090, DOI: 10.35940/invitee.A4144.119119.

[8] H. Snyder, "Literature review as a research methodology: An overview and guidelines," J. Bus. Res., vol. 104, no. 2019, pp. 333-339, DOI: 10.1016/j.jbusres.2019.07.039.

[9] N. S. Selamat, "An Overview of Big Data Usage in Disaster Management," vol. 11, no. 1, pp. 35-40.

[10] R. P. Schumaker and H. Chen, "Textual Analysis of Stock Market Prediction Using Breaking Financial News: The AZFinText System," no. 2006, pp. 1 -29.

[11] Mohammed Ali Shaik and Dhanraj Verma, "Prediction of Heart Disease using Swarm Intelligence based Machine Learning Algorithms", International Conference on Research in Sciences, Engineering & Technology, AIP Conf. Proc. 2418, 020025-1–020025- 9; https://doi.org/10.1063/5.0081719, Published by AIP Publishing. 978-0-7354-4368-6, pp. 020025-1 to 020025-9

[12] M. K. A. Shatnawi, " Stock Price Prediction Using K -Nearest Neighbor (k NN) Algorithm," vol. 3, no. 3, pp. 32-44, 2009.

[13] C. Hargreaves, "Prediction of Stock Performance Using Analytical Techniques," vol. 5, no. 2, pp. 136-142, DOI: 10.4304/jetwi.5.2.136-142.

[14] Magaji, A.S," An intense Nigerian stock exchange market prediction using logistic with back-propagation ANN model", Science World Journal, 9(2), pp.8-13, 2014.

[15] Mohammed Ali Shaik, "Time Series Forecasting using Vector quantization", International Journal of Advanced Science and Technology (IJAST), ISSN:2005-4238,Volume-29,Issue-4 (2020), Pp.169-175.

[16] Shashank, D. S., V. Sruthi, M. L. Vijayalakshimi, and Jacob Shimona Garcia. "Turnover prediction of shares using data mining techniques: a case study." arXiv preprint arXiv:1508.00088,2015.

[17] Mohammed Ali Shaik, "A Survey on Text Classification methods through Machine Learning Methods", International Journal of Control and Automation (IJCA), ISSN:2005-4297,Volume-12,Issue-6 (2019), Pp.390-396.

[18] Khan, W., M. A. Ghazanfar, M. Asam, A. Iqbal, S. Ahmad, and Javed Ali Khan. "Predicting Trend In Stock Market Exchange Using Machine Learning Classifiers." Science International 28, no. 2,2016.

[19] Mohammed Ali Shaik, Dhanraj Verma, P Praveen, K Ranganath and Bonthala Prabhanjan Yadav, (2020), RNN based prediction of spatiotemporal data mining, 2020 IOP Conf. Ser.: Mater. Sci. Eng. 981 022027, doi.org/10.1088/1757-899X/981/2/022027

[20] Zaidi, M., and A. Amirat. "Forecasting stock market trends by logistic regression and neural networks: Evidence from KSA stock market." Int.J. Econ. Commer. Manag 4, vol. IV, no. 6, pp. 220-234, 2016

[21] Mohammed Ali Shaik and Dhanraj Verma, (2020), Deep learning time series to forecast COVID-19 active cases in INDIA: A comparative study, 2020 IOP Conf. Ser.:Mater.Sci.Eng. 981 022041, doi.org/10.1088/1757-899X/981/2/022041

[22] A. Nayak, M. M. M. Pai, and R. M. Pai, "Prediction Models for Indian Stock Market," Procedia - Procedia Comput. Sci., vol. 89, pp. 441-449, 1877, DOI: 10.1016/j.procs.2016.06.096.

[23] H. Karchalkar, A. Jain, A. Singh, and V. Kumar, "Open Price Prediction of Stock Market using Regression Analysis," vol. 6, no. 5, pp. 418-421, 2007, DOI: 10.17148/IJARCCE.2017.6578.

[24] Y. Yaslan, " Stock daily return prediction using expanded features and feature selection," no. 2017, pp. 4829-4840, DOI: 10.3906/elk-1704- 256.

[25] Mohammed Ali Shaik and Dhanraj Verma, (2020), "Enhanced ANN training model to smooth and time series forecast", 2020 IOP Conf. Ser.: Mater. Sci. Eng. 981022038.

[26] Mohammed Ali Shaik, S. k. Koppula, M. Rafiuddin and B. S. Preethi, "COVID-19 Detector Using Deep Learning," 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 2022, pp. 443-449, doi: 10.1109/ICAAIC53929.2022.9792694.

[27] Mohammed Ali Shaik and Dhanraj Verma, "Prediction of Heart Disease using Swarm Intelligence based Machine Learning Algorithms", International Conference on Research in Sciences, Engineering & Technology, AIP Conf. Proc. 2418, 020025-1–020025-9; https://doi.org/10.1063/5.0081719, Published by AIP Publishing. 978-0-7354-4368-6, pp. 020025-1 to 020025-9

[28] Mohammed Ali Shaik and Dhanraj Verma, "Predicting Present Day Mobile Phone Sales using Time Series based Hybrid Prediction Model", International Conference on Research in Sciences, Engineering & Technology, AIP Conf. Proc. 2418, 020073-1–020073-9; https://doi.org/10.1063/5.0081722, Published by AIP Publishing. 978-0-7354-4368-6, pp. 020073-1 to 020073-9

[29] Mohammed Ali Shaik, MD.Riyaz Ahmed, M. Sai Ram and G. Ranadheer Reddy, "Imposing Security in the Video Surveillance", International Conference on Research in Sciences, Engineering & Technology, AIP Conf. Proc. 2418, 020012-1–020012-8; https://doi.org/10.1063/5.0081720, Published by AIP Publishing. 978-0-7354-4368-6, pp. 020012-1 to 020012-8.

[30] Mohammed Ali Shaik, Geetha Manoharan, B Prashanth, NuneAkhil, Anumandla Akash and Thudi Raja Shekhar Reddy,"Prediction of Crop Yield using Machine Learning", International Conference on Research in Sciences, Engineering & Technology, AIP Conf. Proc. 2418, 020072-1–020072-8; https://doi.org/10.1063/5.0081726, Published by AIP Publishing. 978-0-7354-4368-6, pp. 020072-1 to 020072-8