IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-8, Aug 2018 E-ISSN: 2347-2693

Stock Market Close Price Prediction Using Neural Network and Regression Analysis

Prateek Purey[1*], Anil Patidar[2]

Computer Science,Acropolis Institute of Technology and Research,RGPV, Indore, India

Corresponding Author: prateekpurey1000@gmail.com, Tel.: +91-99267-88853

Available online at: www.ijcseonline.org

Accepted: 17/Aug/2018, Published: 31/Aug/2018

***Abstract*—** The financial market is very dynamic in nature and changing continuously. In addition of that due to it's dynamicity prediction of stock price values are not much accurate. In order to predict the stock's close values accurately machine learning technique is used in this proposed work. The proposed technique usages supervised learning technique because supervised learning techniques can predict values more accurately. In order to train and test the proposed machine learning prediction technique the YQL data in offline mode is used. The proposed stock market price prediction method is a hybrid data model. In this context two different algorithms are combined for obtaining the goodness of both the techniques. Here both the algorithms are analyze data according to their methodology and perform prediction. After that both algorithms' approximated values are combined by computing the mean values as final prediction. Therefore the proposed technique optimizes the performance of the traditional back propagation based stock market price prediction. The implementation of the proposed technique is performed using the JAVA technology and the performance of the system is measured in term of accuracy, error rate, time complexity and memory usages. The performance of the system demonstrate the proposed technique enhance the prediction of the close values. In addition of that comparative performance study of proposed technique is performed with the traditional back propagation model using experimental outcomes. Results demonstrate proposed technique out perform with respect to the traditional approach of stock market price prediction.

***Keywords*—** Stock Market Price Prediction, Regression Analysis, Error Adjustment

I. INTRODUCTION

The data mining techniques are now in these days utilized in a number of applications such as in banking, finance, education sectors and others. These applications use the data mining techniques to make accurate analysis of data; by which suitable decisions according to the situations and available information is performed. In this context supervised learning and unsupervised learning techniques can be used. The supervised learning techniques are able to learn on the pattern available on the initial training samples. Using the initial samples trained model generates mathematical systems that are used for prediction and/or classification. The classification is sometimes different as compared to prediction, because the classification techniques predicts the strict class labels defined in the training data instances and prediction can provide the approximate values that are near about the target data. In this presented work the financial market data is tried to predict.

The prediction of the stock market data is very popular domain of data forecasting application. But the approximation of the stock market prices are not much easy directly by using the algorithms because the stock market prices are rapidly changing according to the various conditions for a country such as changes in company, contracts of the company, political changes and others. Therefore by using only single algorithm based data analysis it is not much accurate. In this presented work the stock market prediction technique is performed using a hybrid algorithm that algorithm is combination of the linear regression and back propagation neural network (BPN). Therefore prediction performed with the single data model is enhanced through the combining the output of two different prediction techniques.

II. RELATED WORK

The data mining and machine learning techniques are used for classification, prediction or association rule analysis. The data mining technique are usages the historical data and

trends to understand the patterns of the data. Additionally after understanding or learning with the initial provided samples, the system is able to find or recognize the similar trend in unlabelled data. The applicability of algorithms are depends on the structure or formats of the data. The proposed technique is implementation of a supervised learning algorithm. The aim of the proposed system modeling is to enhance the traditional stock market price prediction technique using the neural network and regression analysis. Therefore a hybrid model for predicting the stock market price is proposed in this work. The proposed technique usages the goodness of the linear regression based learning and back propagation neural network based technique, for enhancing the prediction accuracy of model. The technique usages the YQL stock price data for training and testing purpose. The YQL data base is basically provided by the Yahoo Corporation and can be extracted using the date range and company name. This database allows us to search the stock market prices for the particular company name. The data here used in the offline format therefore after extraction of data the CSV file formats are used to store the stock market prices according to company names. After the training sample collection both the models are trained separately and the prediction outcomes are combined using the mean values of the predictor. This section provides the overview of the proposed work in next section the detailed analysis and design of the system is explained.

III. METHODOLOGY

The proposed technique for stock market price prediction is demonstrated the figure 2.1. This diagram contains the entire process of the training and testing of the proposed prediction scheme.

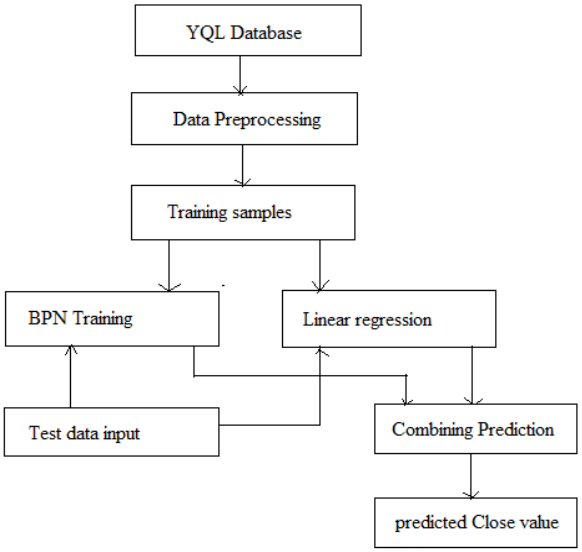


Fig.1- System Architecture

YQL data: that is Yahoo financial data base based data collection. The Yahoo provides an API by which we can extract the financial price history of a target company according to the given date range and the company code. This data is downloaded and utilized as the historical trend of particular market price trend. The available data in YQL is in structured format.

Data preprocessing: the aim of preprocessing is to clean and transform the data. By which good quality data can be used with the learning algorithm. Here the data is available in different range of values in the given attributes of dataset. Therefore a data scaling or normalization technique is used for preprocessing of data. To normalize the data and scale them in a particular range the min max technique is used. Therefore first the maximum and minimum value is computed and using these maximum and minimum values the scaling of each value is performed. For scaling of values the following formula is used:

$$scaled\ value = \frac{current\ value - minimum\ value}{maximum\ value - minimum\ value}$$

Training samples: the normalized values are used with the both data models for training and testing purpose.

Regression analysis: here for analyzing the stock market data the linear regression technique is used. Regression analysis is used to find equations that fit data. Once we have the regression equation, we can use the model to make predictions. One type of regression analysis is linear analysis. When a correlation coefficient shows that data is likely to be able to predict future outcomes and a scatter plot of the data appears to form a straight line, you can use simple linear regression to find a predictive function. If you recall from elementary algebra, the equation for a line is y = mx + b. To calculate linear regression, and find the equation y' = a + bx.

Linear regression is a way to model the relationship between two variables. You might also recognize the equation as the slope formula. The equation has the form Y=a+bX, where Y is the dependent variable (that's the variable that goes on the Y axis), X is the independent variable (i.e. it is plotted on the X axis), b is the slope of the line and a is the y-intercept.

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x^2)}$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x^2)}$$

Values adjustment: the error difference is computed and adjusted in this phase. Therefore the initial predicted value is adjusted using the computed sentiment score category. That works as the factor of adjustment. To understand the value adjustment we consider the following example:

Let the initial prediction of the data is P_i and the actual value of stock price is A_i. Therefore the error is

$$E = P_i - A_i$$

Now the previous predicted values error is involve with the current value of prediction S_i therefore the value adjustment is:

$$Adj = E * S_i$$

BPN training: The implementation of neural network is defined in two phases' first training and second prediction: training method utilizes data and designs the data model. By this data model next phase prediction of values is performed.

Training:

1. Prepare two arrays, one is input and hidden unit and the second is output unit.
2. Here first is a two dimensional array W_{ij} is used and output is a one dimensional array Y_i.
3. Original weights are random values put inside the arrays after that the output is given as**.**

$$x_j = \sum_{i=0} y_i W_{ij}$$

Where, y_i is the activity level of the j^{th} unit in the previous layer and W_{ij} is the weightof the connection between the i^{th} and the j^{th} unit.

4. Next, action level of y_i is estimated by sigmoidal function of the total weighted input.

$$y_i = \left[\frac{e^x - e^{-x}}{e^x + e^{-x}}\right]$$

When event of the all output units have been determined, the network calculates the error (E) given in equation**.**

$$E = \frac{1}{2}\sum_i (y_i - d_i)^2$$

Where, y_i is the event level of the j^{th} unit in the top layer and d_i is the preferred output of the j_i unit.

Calculation of error for the back propagation algorithm is as follows:

- Error Derivative (EA_j) is the modification among the real and desired target:

$$EA_j = \frac{\partial E}{\partial y_j} = y_j - d_j$$

- Error Variations is total input received by an output changed

$$EI_j = \frac{\partial E}{\partial X_j} = \frac{\partial E}{\partial y_j} X \frac{dy_j}{dx_j} = EA_j y_j (1 - y_i)$$

- In Error Fluctuations calculation connection into output unit is required:

$$EW_{ij} = \frac{\partial E}{\partial W_{ij}} = \frac{\partial E}{\partial X_j} = \frac{\partial X_j}{\partial W_{ij}} = EI_j y_i$$

- Overall Influence of the error:

$$EA_i = \frac{\partial E}{\partial y_i} = \sum_j \frac{\partial E}{\partial x_j} X \frac{\partial x_j}{\partial y_i} = \sum_j EI_j W_{ij}$$

Test set: in order to test the trained hybrid data model the test values on the system is produced the system analyzes the data according to their learning and produces the close value.

Combining the results: the system is combines the output of both the model. Therefore the predicted value of the neural network and regression based analysis is combined using the following formula:

$$final\ prediction = \frac{prediction\ of\ BPN + prediction\ of\ Regression}{2}$$

Using this formula the next step values are predicted.

IV. RESULTS AND DISCUSSION

The given chapter provides the results development and the performance analysis of the proposed and base algorithm. Therefore to compare the performance essential performance factors are evaluated and their results are reported.

A. Accuracy

The accuracy of a machine learning system is the unit of closeness of a quantity to that required to be predict in other words the amount of correctly classified samples that are need to be predict is termed as machine learning system accuracy. The accuracy of the system is computed using the following formula:

$$\text{Accuracy} = \frac{\text{Total correctly identified patterns}}{\text{Total Patterns to classify}} \text{X100}$$

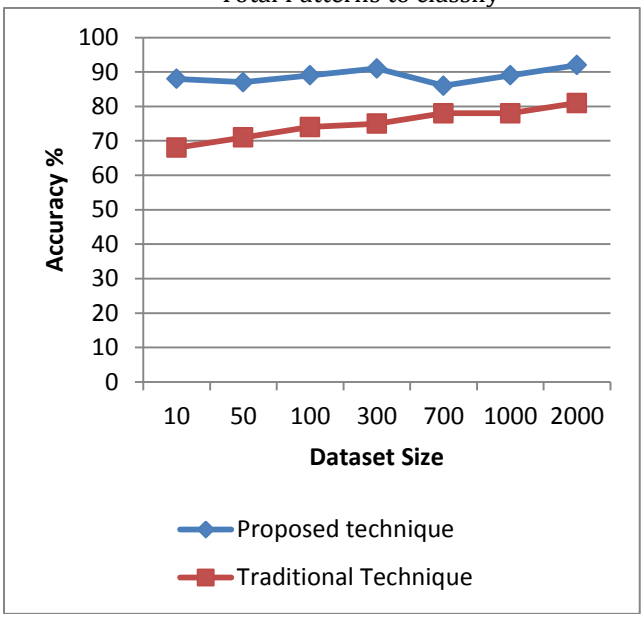


Figure 3.1 Accuracy

The figure 3.1 shows the comparison of the proposed and base system the proposed system is hybrid model based on regression analysis and back propagation neural network and the traditional model is simple back propagation neural network. Additionally the obtained performance data is also given in table 3.1. In this diagram X-axis contains the amount of data instances for learning and Y-axis shows

accuracy in terms of percentage. In order to represent the performance of both the implemented prediction models the red line used to demonstrate the performance of classical approach and the blue line shows the performance of the proposed approach. According to the graph and table results the proposed technique provides high accurate results as compared to traditional approach.

Table 3.1 Accuracy Performance

[1] **Dataset Size**	[2] **Proposed Method**	[3] **Base Method**
[4] 10	[5] 88	[6] 68
[7] 50	[8] 87	[9] 71
[10] 100	[11] 89	[12] 74
[13] 300	[14] 91	[15] 75
[16] 700	[17] 86	[18] 78
[19] 1000	[20] 89	[21] 78
[22] 2000	[23] 92	[24] 81

B. Error Rate

Error rate is the measurement of the performance in terms of misclassification rate. In other words the amount of data which is not correctly recognized using the trained classifier is termed as the error rate of the system. That can be evaluated using the following formula:

$$\text{Error Rate} = 100 - \text{Accuracy}$$

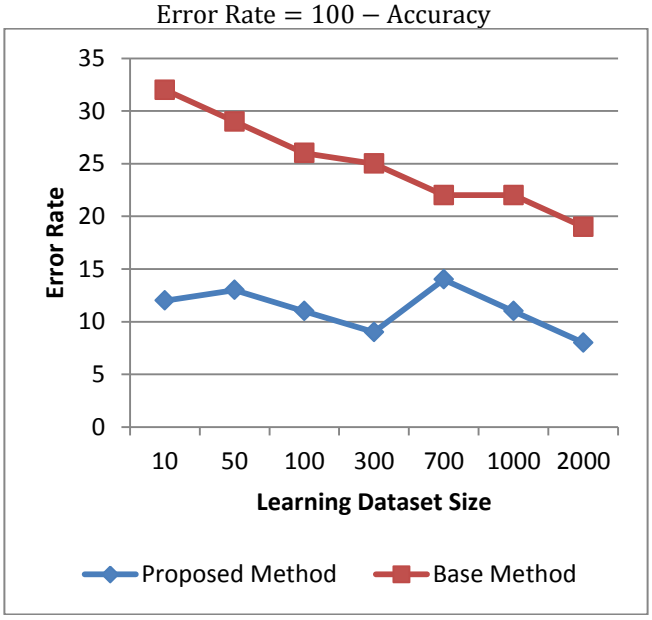


Figure 3.2 Error Rate

The graph 3.2 represents the Error rate of implemented proposed regression analysis based neural network and traditional artificial neural network. In this graph the blue line represents the error rate of the proposed technique and red lines of traditional algorithms. The X-axis of the given graph demonstrates the data base instances used for training of the system and the Y-axis shows how much error rates are obtained by implement to both scenarios. According to the obtained result error in percentage is very fewer to base neural method and the proposed system delivers the minimization of data error.

Table 3.2 Table Error Rate

[25] **Data set Size**	[26] **Proposed Method**	[27] **Base Method**
[28] 10	[29] 12	[30] 32
[31] 50	[32] 13	[33] 29
[34] 100	[35] 11	[36] 26
[37] 300	[38] 9	[39] 25
[40] 700	[41] 14	[42] 22
[43] 1000	[44] 11	[45] 22
[46] 2000	[47] 8	[48] 19

C. Time Consumption

The amount of time required to learn the given patterns from input training samples using the implemented prediction algorithm is known as the time consumption or time complexity of the implemented algorithms. The figure 3.3 and table 3.3 contains the time consumption of the base algorithm and the proposed BPN based algorithm in terms of seconds. To represent the performance of the classifier the proposed technique is demonstrated using the red line graph and the blue line shows the performance of the traditional approach of learning. According to the given results the proposed technique consume less time for training of input of stock market as compared to the base method. Therefore the proposed system is less time consuming for prediction of the data patterns.

Table 3.3 Time Consumption

[49] **Dataset Size**	[50] **Proposed Method**	[51] **Base Method**
[52] 10	[53] 55	[54] 62
[55] 50	[56] 61	[57] 71
[58] 100	[59] 69	[60] 88
[61] 300	[62] 74	[63] 91
[64] 700	[65] 81	[66] 97
[67] 1000	[68] 94	[69] 106
[70] 2000	[71] 99	[72] 120

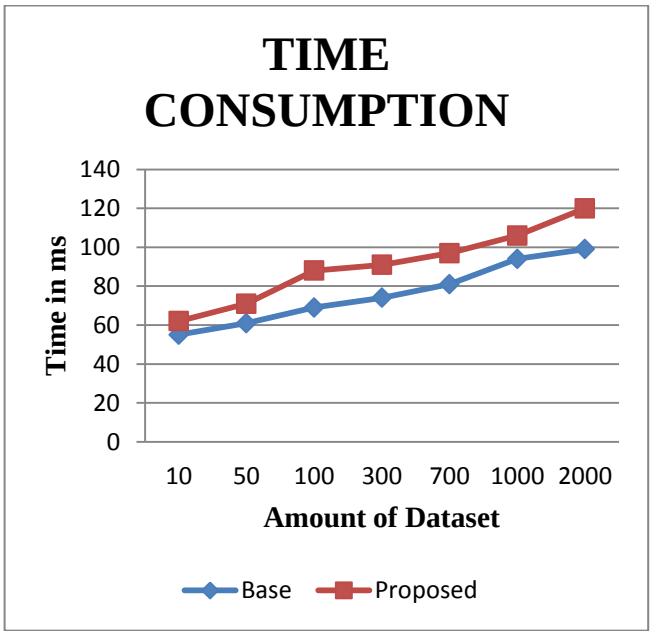


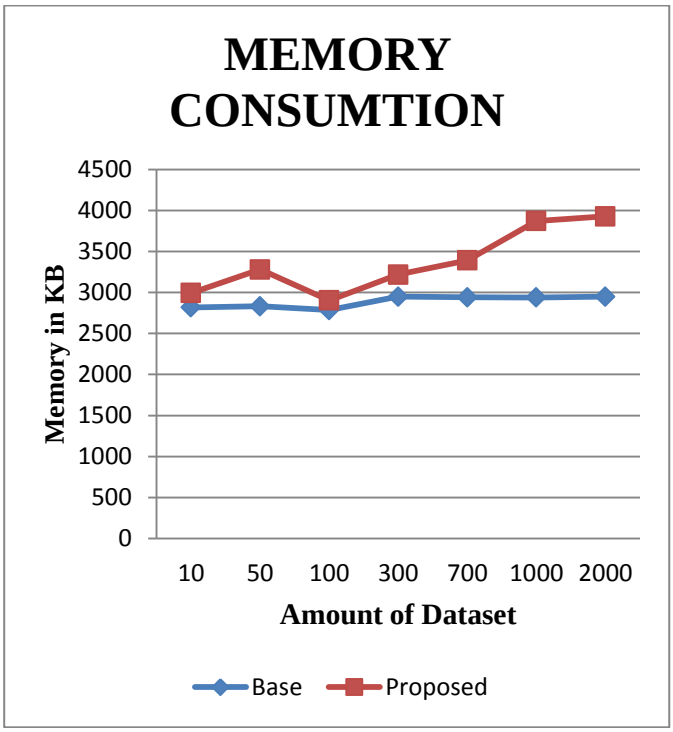
Figure 3.3 Time Consumption

D. Memory Consumption

The quantity of main memory required to implement the algorithm is termed as the space complexity of the system. That is sometimes also called the memory consumption of the algorithms. The comparative memory consumption of both the algorithms namely Base method and Artificial Neural network based Proposed Algorithm is given using figure 3.4 also data shown in table 3.4. In this diagram the X-axis includes amount of dataset stored in database, and the Y-axis shows the amount of memory consumed during processing of the data. The given memory consumption is provided here in terms of kilobytes. According to the evaluated results the performance of the traditional technique is less efficient as long as slow performance as compared to the proposed method.

Table 3.4 Memory Consumption

[73] **Dataset Size**	[74] **Proposed Method**	[75] **Base Method**
[76] 10	[77] 2817	[78] 2994
[79] 50	[80] 2831	[81] 3281
[82] 100	[83] 2784	[84] 2904
[85] 300	[86] 2948	[87] 3217
[88] 700	[89] 2940	[90] 3392
[91] 1000	[92] 2938	[93] 3872
[94] 2000	[95] 2948	[96] 3928



.

V. CONCLUSION AND FUTURE SCOPE

Conclusion:- Data mining and machine learning techniques enable us for analyzing the hidden trends on the data and recognize the similar patterns or trends in newly appeared data. Therefore the data mining techniques are used in different applications for making decisions and forecasting the values. In this presented work the use of data mining technique is performed for prediction of stock market price close values. There are a number of data models are available that claim to produce the accurate close prices for the stock market but most of the techniques are either time consuming or expensive in terms of resource consumption. Therefore a lightweight model is introduced in this work for efficiently prediction of the stock market price.

In order to develop the stock market price prediction model the regression analysis technique and neural network based technique is combined for improving the predicted prices of the proposed algorithm. The data attributes is basically available in different scales therefore the data is normalized and categorized for finding the trends of value change of the target stock price. The YQL database is used as the training sample or the historical data trend for learning with the supervised learning algorithms. The regression technique is most of time used to understand the historical trends of the data and predict the future or one step ahead data. Therefore YQL data is analyzed with the regression algorithm and BPN algorithm for new values prediction.

The implementation of the proposed technique is provided using the JAVA technology. Additionally the YQL

dataset is used for consuming with the prediction model. The experiments with the implemented algorithm is conducted and based on the experiments observations the summary of performance is reported using table 4.1

Table 4.1 Performance Summary

[97] S. No.	[98] Parameters	[99] Proposed technique	[100]Traditional technique
[101] 1	[102] Accuracy	[103] High	[104] Low
[105] 2	[106] Error rate	[107] Low	[108] High
[109] 3	[110] Time consumption	[111] Low	[112] High
[113] 4	[114] Memory consumption	[115] Low	[116] High

The proposed data model for stock market price prediction is accurate and also efficient in terms of time and memory resource consumption. Therefore the proposed model is acceptable for the real world application use.

Future scope:- The main aim of the proposed work is to enhance the traditional stock market price prediction technique is developed successfully. The implemented system provides the accurate results therefore the proposed technique is acceptable but the following extension is proposed for extending the given work for more accurate prediction in different situations:

- ✓ Include the different kinds of data sources according to the situation for finding the values adjustment on the error
- ✓ Regression analysis can be predict the values but the nature of market is very fluctuating therefore in near future deep learning technique is need to be implement the price prediction model

REFERENCES

[1] Moghaddam, Amin Hedayati, Moein Hedayati Moghaddam, and Morteza Esfandyari, "Stock market index prediction using artificial neural network", Journal of Economics, Finance and Administrative Science 21, no. 41 (2016), pp. 89-93.

[2] Gorunescu, F, Data Mining: Concepts, Models, and Techniques, Springer, 2011.

[3] Han, J., and Kamber, M., Data mining: Concepts and techniques, Morgan-Kaufman Series of Data Management Systems San Diego: Academic Press, 2001.

[4] Neelam adhab Padhy, Dr. Pragnyaban Mishra and Rasmita Panigrahi, "The Survey of Data Mining Applications and Feature Scope, International Journal of Computer Science, Engineering and Information Technology (IJCSEIT)", vol.2, no.3, June

[5] Introduction to Data Mining and Knowledge Discovery, Dunham, M. H., Sridhar, S., "Data Mining: Introductory and Advanced Topics", Pearson Education, New Delhi, 1st Edition, 2006.

[6] Mohammed J. Zaki and Wagner Meira Jr, "Data Mining and Analysis Fundamental Concepts and Algorithms", Cambridge University Press Hardback, 2014 [Book]

[7] Neelamadhab Padhy, Dr. Pragnyaban Mishra, "The Survey of Data Mining Applications and Feature Scope", International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), PP. 43-58 Vol.2, No.3, June 2012.

[8] Tao Li, Steve Luis, Shu-Ching Chen, Vagelis Hristidis. "Using data mining techniques to address critical information exchange needs in disaster affected public-private networks", Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2010.

[9] Kantardzic, Mehmed ―Data mining: Concepts, Models, Methods, and Algorithms, John Wiley & Sons, 2003.

[10] Ian H. Witten; Eibe Frank; Mark A. Hall, ―Data Mining: Practical Machine Learning Tools and Techniques (3rd Ed.), Elsevier, 30 January 2011.

[11] Shalev-Shwartz, Shai, and Shai Ben-David. Understanding machine learning: From theory to algorithms, Cambridge University press, 2014.c.

Authors Profile

Mr.Prateek Purey is pursing M.Tech Computer Science and Engineering in the Department of Computer Science and Engineering, Acropolis Institute of Research & Technology, Indore. he received her Bachelor of Engineering Degree from Shree Venkteshwar Institute of Technology, Indore. Her research interests are Data mining, Big data Analytics and Network Security.

Mr Anil patidar pursed Bachelor of Engineering in Computer Science and Master of Technology from school of Information Technology UGC RGPV Bhopal Madhy Pradesh India, He is currently pursuing Ph.D. from DAVV Indore and currently working as Assistant Professor in Department of Computer Science Engineering in Acropolis Institute of Research and Technology.His main research work focuses on Network Security, Cloud Security and Privacy, Big Data Analytics, Data Mining, Block chain,softwere engg. He has 9 years of teaching experience and 4 years of Research Experience.

