

# Forecasting Techniques for Sales Prediction

V.Phanindra Reddy, N.Lokesh Kumar, P.Krishna Kanth, P.S.V.S.Sridhar

**Abstract:** All of us are very curious about future, very excited to know what will happen in the next moment. Similarly, retailers are also curious about the future of their business, its progress and their future sales. Walmart is the world's biggest retailer and also has a vast grocery chain over the world. It was initially established in America 1962. In 2019, it has more than 11,000 stores in 28 countries but the sales differ from place to place. Many sales strategies, discount rates will be introduced for the improvement of sales. Retailers always try to attract the common people to visit their store. They always focus on improving the future sales. Using some Machine learning forecasting models, we can estimate the future sales based on the past data. Our aim is to apply time series forecasting models to retail sales data, which contains weekly sales of 45 Walmart stores across United States from 2010 to 2012. There are other factors which effects the analysis of weekly sales - markdown, consumer per index, Is Holiday (boolean value returns whether it is holiday or not), size of the store, unemployment, store type, fuel price and temperature. The forecasting models applied for the data are Autoregressive Integrated Moving Average (ARIMA) model and Feed Forward Neural Networks (FFNN). The dataset will be divided into training and testing datasets. The predicted values will be checked with the test data and accuracy will be calculated. Based on the accuracy we conclude which of the two models will better for the sales prediction.

**Keywords :** ARIMA, FFNN, ML, AI.

## I. INTRODUCTION

This Many research studies proposed various predictions on the stock market and proposed various solutions. The noticeable methods fall into two general classifications, in particular, measurable and delicate processing procedures. Measurable procedures incorporate, among others, exponential smoothing, and Auto Regressive Integrated Moving Average (ARIMA), and summed up autoregressive conditional heteroskedasticity (GARCH) volatility [1]. The proposed model is also called a Box-Jenkins model or system, is usually utilized in analysis and weather forecasts. It is generally viewed as the most proficient anticipating procedure in sociology and is utilized widely for time arrangement. The utilization of ARIMA for anticipating time

Revised Manuscript Received on January 27, 2020.

**V.Phanindra Reddy**, student, Department of computer science and engineering, K L Educational foundation, Deemed to be University, Vaddeswaram, Andhra Pradesh, India.

**N.Lokesh Kumar**, student, Department of computer science and engineering, K L Educational foundation, Deemed to be University, Vaddeswaram, Andhra Pradesh, India.

**P.Krishna Kanth**, Student, Department of computer science and engineering, K L Educational foundation, Deemed to be University, Vaddeswaram, Andhra Pradesh, India.

**P.S.V.S.Sridhar**, Associate Professor, Department of Computer Science and Engineering, K L Educational foundation, Vaddeswaram, Andhra Pradesh, India.

arrangement is basic with vulnerability as it doesn't expect information on any basic model or connections as in some different techniques. ARIMA basically depends on past estimations of the arrangement just as past blunder terms for determining [2, 3].

Its wide utilization is because of the few distinctive highlights of ANNs that make them alluring to the two specialists and modern professionals. As expressed in [4], ANNs

are information-driven, self-versatile techniques with not many earlier suspicions. They are likewise great indicators with the capacity to summed up objective facts from the outcomes gained from unique information, along these lines allowing right surmising of the inert piece of the populace. Besides, ANNs are general approximate as a system can effectively inexact a ceaseless capacity to the ideal degree of exactness. At long last, ANNs have been seen as exceptionally proficient in taking care of nonlinear issues incorporating those in the genuine world [4]. This is rather than numerous conventional methods for time arrangement expectations, for example, ARIMA, which accepts that the arrangement is produced from direct procedures and accordingly may be wrong for most certifiable issues that are nonlinear [5, 6]. There is a developing a need to settle profoundly nonlinear, time-variation issues the same number of uses [7].

In this paper, the proposed system is to predict weekly sales of the Walmart based on historical data provided. Analyse the multi variant data and plot the graphs of the variables with the target variable. Normalize the data using min max normalization. Apply STL Decomposition to the data and observe the behaviour of the curve over the period of time.

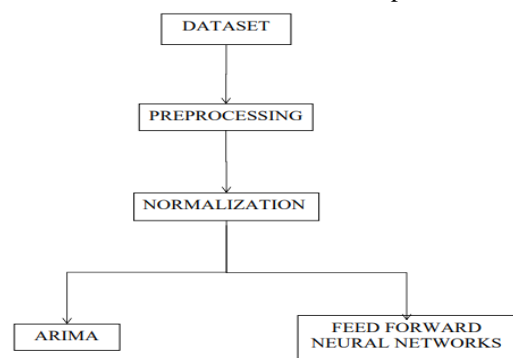


Fig. 1.Procedure of Machine

II. LITERATURE SURVEY

In this chapter two models are discussed such as ARIMA is the first model and ANN is the second model.

The improvement is calculated by the two models among that second model, and genetic algorithms (GAs) in displaying time arrangement info captivated with demonstrating exactness, comfort, and procedure time [9].

The examination uncovered that direct advancement methods gave the most effective gauges with GAs giving comparable outcomes if the bounds of the parameters and also the goals were deliberately chosen, whereas NNs gave the foremost passing awful gauges.

The work described in [10] in addition analyzed the anticipating execution of the above discussed models in estimating the Korean Stock indicant. The first model usually gave a lot of precise conjectures than the back-engendering neural system model utilized. This can be more and more articulated for the midrange forecasting skylines.

These models accomplished nice forecast execution in some certifiable applications notably time arrangement expectation. Trial results got by the creators any uncovered that first model for the foremost half performs higher within the expectation of direct time arrangement, whereas second perform higher within the forecast of nonlinear time arrangement. During a comparative report for financial determinant disclosed in, second model was seemed to perform superior to first model in forecasting, whereas first model performed superior to second in directional estimating.

The stock determinant execution of ANN and ARIMA models and incontestible that the ANN model noninheritable most popular returns over the regular ARIMA models equally [14], The expectation execution of ANNs and ARIMA on time arrangement forecast to point out that the ANNs beat ARIMA in foreseeing stock development heading because the last had the choice to acknowledge hid examples within the info utilized [15].

The forecasting execution of first and second models in estimating day by day most extreme gas focus. Experimental results got likewise indicated that the second model is healthier than the first model [16]. The comparative correlation captivated with the state stock trade and showed signs of improvement accuracy with ANN than the ARIMA model [17]. a lot of writing has indicated the pervasive utilization of ANNs as robust equipment for the stock worth forecast. This makes ANN a promising strategy or potential 0.5 and 0.5 for the expectation of development in time arrangement.

III. DATASET DESCRIPTION.

The dataset contain historical weekly sales data of 45 Walmart stores across United States from 2010 to 2012. It has a total of 423,326 records. Each store has 98 to 99 departments. Each sample has following values: Store: The store number Department: The department number Date: Weekly date Weekly Sales: Sales of a given department of a particular store. IsHoliday: Whether the week has a special holiday Temperature: Average Temperature in that region Fuel Price: cost of fuel in the region Markdown (1-5): Pricing

strategy on items CPI (Consumer per index): measures changes in price level of goods Unemployment: The Unemployment rate Store Type: the type of the store (A, B or C) Size: Size of the store is based on the number of items in the store.

IV. ARIMA

It is a popular and widely used statistical method for time series forecasting. ARIMA models are time series regression models. In regression we have a dependent variable Y and independent variable X and we linearly regress Y on X. The errors are white noise.

$$Y_i = \beta X_i + \epsilon_i$$

Auto Regression (AR): An auto regressive model predicts future behaviour based on past behaviour. It is used for time series when there is correlation between values in a time series. With time series we can regress today on yesterday. What happens today is the dependent variable and yesterday is the independent variable.

$$X_t = \phi X_{t-1} + \epsilon_t \quad \epsilon_t \text{ is white noise}$$

Moving Average (MA): Typically, time series data is correlated and assuming the errors are not correlated may lead to bad forecasts. One way to overcome this problem is to use moving average model. In the below formula, the error of time t is correlated with the error of time t-1.

$$\epsilon_t = W_t + \theta W_{t-1}$$

ARMA (Auto Regressive Moving Average) : Putting the Auto Regression and the Moving Average together leads to ARMA model. This is also called Auto regression and auto correlated errors

$$X_t = \phi X_{t-1} + W_t + \theta W_{t-1}$$

V. FEED FORWARD NEURAL NETWORKS (FFNN):

Neural network is a computational system that predicts the existing data. Construction of neural network Input layers: Takes inputs from existing data Hidden layers: uses back propagation to optimize the weights of the input variables in to improve the prediction of the model Output layers: Output of predictions depends on the data from the input layers and hidden layers

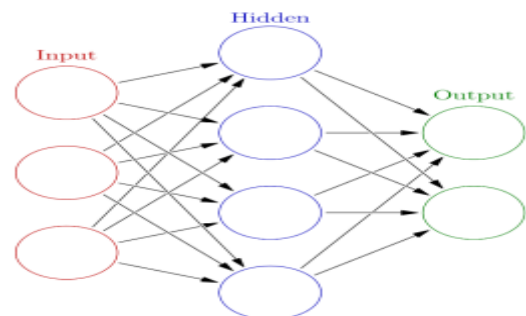


Fig. 2.Examples of Neural Networks



Training and Test Data  
Training Data: 2010 to 2012  
Testing Data: 2013

### VI. NORMALIZATION

Min Max Normalization is used to normalize or standardize the data. As our data is bivariate and the values are in thousands, the algorithm may not process the data legitimately. So we need to normalize the data so that all the values will be between 0 and 1.

$$v' = \frac{v - \min_F}{\max_F - \min_F}$$

### VII. TRAINING NEURAL NETWORK MODEL

The limit is set as 0.01, so that on the off chance that the adjustment in mistake while a cycle is under 1%, at that point no further advancement will be finished. The straight yield variable is set as "bogus", given the effect of the autonomous factors on the needy variable (profit) is thought to be non-direct. Expanding the number of shrouded hubs has more odds of expanding precision. We can create the mistake of the neural system model, incorporating the loads between the information sources, shrouded layers, and results.

#### Testing the Accuracy:

The neural network has been created with the training dataset. We need to contrast preparing information with the test information to get the exactness of the neural system conjecture. "Subset" work wipes out the needy factors from the test information "Register" work then forecast variable then the outcomes contrasts the anticipated information and the real information Confusion Matrix is utilized for accuracy.

### VIII. EVOLUTION RESULTS

Average Sales for all the stores

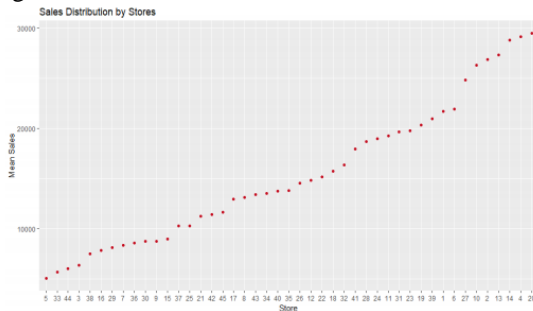


Fig. 3. Average sale of all stores

Average Weekly Sales vs Is Holiday: Weekly sales are more when there is a special holiday in that week.

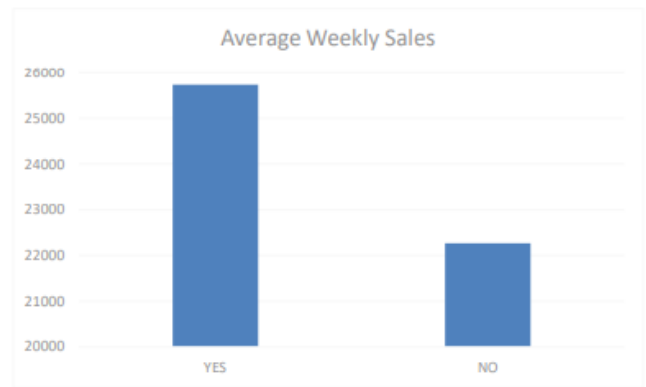


Fig. 4. Average weekly sales

### IX. PREDICTION FOR ARIMA MODEL

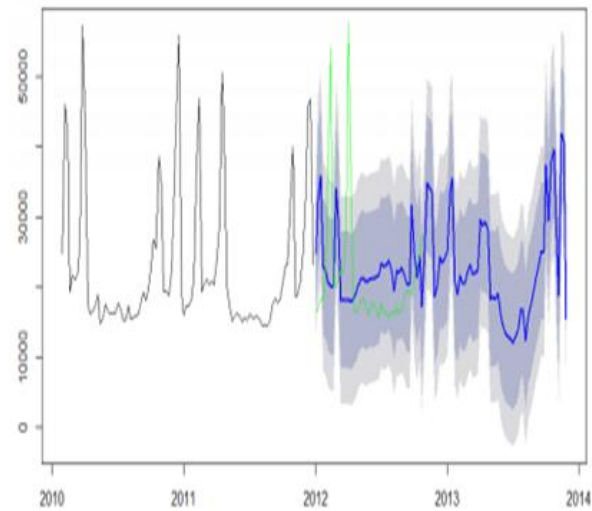


Fig. 5. Prediction for ARIMA model

### X. RESULTS FOR ARIMA MODEL

Date	Point.Forecast	Actual	Date	Prediction	Actual
06-01-2012	25948.59718	16567.69	01-06-2012	19205.87727	16065.49
13-01-2012	33057.32665	16894.4	08-06-2012	19350.56493	17666
20-01-2012	37598.82719	18365.1	15-06-2012	19144.21376	17558.82
27-01-2012	24558.83235	18378.16	22-06-2012	19424.63896	16633.41
03-02-2012	24053.61245	23510.49	29-06-2012	19526.82713	15722.82
10-02-2012	22189.68962	36988.49	06-07-2012	19095.08996	17823.37
17-02-2012	21943.07046	54060.1	13-07-2012	20997.70831	16566.18
24-02-2012	22186.12592	20124.22	20-07-2012	21244.18348	16348.06
02-03-2012	34193.30842	20113.03	27-07-2012	21786.97908	15731.18
09-03-2012	32827.38128	21140.07	03-08-2012	20378.53617	16628.31
16-03-2012	21143.97997	22366.88	10-08-2012	16364.75267	16119.92
23-03-2012	18685.99148	22107.7	17-08-2012	20200.6819	17330.7
30-03-2012	18403.46704	28952.86	24-08-2012	20088.52613	16286.4
06-04-2012	19864.30269	57592.12	31-08-2012	23555.01994	16680.24
13-04-2012	19390.73902	34684.21	07-09-2012	23827.6328	18322.37
20-04-2012	17503.25054	16976.19	14-09-2012	22708.68307	19616.22
27-04-2012	17916.25793	16347.6	21-09-2012	23034.79281	19251.5
04-05-2012	18724.63954	17147.44	28-09-2012	26033.2943	18947.81
11-05-2012	19452.82008	18164.2	05-10-2012	24328.87982	21904.47
18-05-2012	19222.22632	18517.79	12-10-2012	23888.94266	22764.01
25-05-2012	18899.00039	16963.55	19-10-2012	24503.389	24185.27
			26-10-2012	19081.72218	27390.81



**XI. FEED FORWARD NEURAL NETWORKS RESULTS**

```

> results
  actual prediction
3  0.62845981 6.279892e-01
7  0.17650248 1.985715e-01
8  0.27155749 2.691426e-01
12 0.03734733 8.485707e-02
13 0.04686452 7.512588e-02
18 0.07015997 6.205696e-02
19 0.04878091 7.380426e-02
22 0.04170899 5.690140e-02
23 0.07319494 6.439163e-02
24 0.06070108 3.434063e-02
28 0.02320371 4.352294e-03
30 0.02918377 1.193194e-02
31 0.03958704 6.443199e-02
32 0.08494696 6.574321e-06
36 0.20556756 1.315421e-01
39 0.55983509 5.817261e-01
41 0.11641038 8.838844e-02
44 0.18534982 7.142791e-02
46 0.70550845 7.245909e-01
47 0.96142377 7.822782e-01
50 0.06555212 1.475843e-01
52 0.09113536 1.283068e-01
55 0.75040501 6.873283e-01
59 0.13463927 1.984653e-01
62 0.21684994 1.233149e-01
64 0.83551617 6.850148e-01
65 0.62652832 6.425252e-01
68 0.01392204 5.356863e-02
69 0.02796974 1.480331e-02
72 0.01023627 4.000930e-03
74 0.01918789 1.503257e-02
75 0.03742909 1.613108e-02
    
```

**XII. CONCLUSION**

Feed Forward Neural Networks was conformed to be very effective model for the sales data. This shows that FFNN can effectively model time series data without any pre-processing such as detrending or deseasonalizing. We can also infer that using 9 hidden units has given the best accuracy. ARIMA has also shown a decent accuracy but it can be improved. ARIMA couldn't forecast for moving holidays such as Easter, Good Friday. Future work will be for improving accuracy for the bivariate data. Improving the dependencies on the target class so that other variables also affects the weekly sales. STL + ARIMA model should be developed for moving holidays. Applying the models for other stores and departments and improve the accuracy.

**REFERENCES**

1. J. J. Wang, J. Z. Wang, Z. G. Zhang, and S. P. Guo, "Stock index forecasting based on a hybrid model", *Omega*, vol. 40, no. 6, pp. 758–766, 2012.
2. B. G. Tabachnick and L. S. Fidell, "Using Multivariate Statistics", Pearson Education, NJ, USA, 4th edition, 2001.
3. Meyler, G. Kenny, and T. Quinn, "Forecasting Irish Inflation Using ARIMA Models," Technical Paper 3/RT/1998, CBIRD, 1998.
4. M. Khashei and M. Bijari, "An artificial neural network (p, d, q) model for time series forecasting," *ESA*, vol. 37, no. 1, pp. 479–489, 2010.
5. G. Zhang, B. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: the state of the art," *IJF*, vol. 14, no. 1, pp. 35–62, 1998.
6. M. Khashei, M. Bijari, and G. A. R. Ardali, "Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (ANNs)," *NC*, vol. 72, no. 4–6, pp. 956–967, 2009.
7. R. Fuller, *Neural Fuzzy System*, Abo Akademic University, 1995.
8. E. Khan, "Neural fuzzy based intelligent systems and applications," pp. 107–139, CRC Press, New York, NY, USA, 2000.

9. N. Tansel, S. Y. Yang, G. Venkataraman, A. Sasirathsiri, W. Y. Bao, and N. Mahendrakar, "Modeling time series data by using neural networks and genetic algorithms," vol. 9, pp. 1055–1060, ASME Press, New York, NY, USA, 1999.
10. K. Lee, Y. Sehwan, and J. Jongdae, "Neural network model versus SARIMA model in forecasting Korean stock price index (KOSPI)," *IIS*, vol. 8, no. 2, pp. 372–378, 2007.

**AUTHORS PROFILE**



**V. Phanindra Reddy**, is a student at the department of computer science and engineering at K L Educational foundation, Deemed to be University, Vaddeswaram, Andhra Pradesh. He is doing his research work in Software Engineering(SE).



**N. Lokesh Kumar**, is a student at the department of computer science and engineering at K L Educational foundation, Deemed to be University, Vaddeswaram, Andhra Pradesh. He is doing his research work in Software Engineering(SE).



**P. Krishna Kanth**, is a student at the department of computer science and engineering at K L Educational foundation, Deemed to be University, Vaddeswaram, Andhra Pradesh. He is doing his research work in Software Engineering(SE).



**P. S. V. S. Sridhar**, is a Associate Professor in Department of Computer Science and Engineering at K L Educational foundation, Vaddeswaram, Andhra Pradesh.