A COMPARATIVE STUDY OF FORECASTING DIFFERENT DATA-SETS BY FEED FORWARD ARTIFICIAL NEURAL NETWORK USING DATA MINING

Meenu Madan, Dr. Sarsij Tripathi

Department of Computer Science & Engineering
IFTM University,
Moradabad

***Abstract*—Forecasting time series data of different fields like finance, production, industries, etc. is an important problem that is receiving great attention of the researchers. Prediction of stock market opening, closing, high, low, etc. are an issue of interest to financial markets. Rice production in country is of great concern. Milk production is also one of the fields which are of concern as it is the basic necessity for a common man. Many prediction techniques have been used in stock forecasting, Rice prediction, milk prediction. Neural networks are one of the most suitable techniques for forecasting the future value considering the past data.**

This paper aims at collecting the past voluminous data belonging to different fields to forecast the future values of the same using Feed Forward Artificial Neural Network along with data mining concepts. We have calculated the Root Mean Square Error between the actual data and the predicted data. We have used Resilient algorithm in this research work. In this research work we have implemented different ANN architecture using Encog Framework supported by Visual Studio to forecast voluminous data of different fields such as stock market, crop production, etc .

***Keywords* -** Prediction, Data Mining, FFANN, Forecasting, Stock Market.

I. INTRODUCTION

A **stock market** is the aggregation of buyers and sellers of shares; these are securities listed on a stock exchange as well as those only traded privately [1].

A stock exchange is a place to trade shares. Companies may want to get their stock listed on a stock exchange. Stock exchange is an organized, managed and regulated financial market where securities are bought and sold at prices governed by the forces of demand and supply [2]. The **stock market** is one of the most important sources for companies to raise money and reputation of the securities the offer in the market. This allows businesses to be publicly traded, or raise additional financial capital for expansion by selling shares of ownership of the company in a public market [1].

Rice is one of the main food grains among the crops produced in India. India is the second largest producer of rice in the world. Due to the increase in population there is an increase in the demand of rice. India is an agricultural country and there is a great uncertainty about the output of the crop in the year. So forecasting the rice production is important for efficient planning of rice production in the country [3][4].

India is predominantly an agrarian society where animal husbandry forms the backbone of national economy. Dairying provides millions of small marginal farmers and landless labors means for their subsistence. On global basis India is able to produce milk at very competitive prices by virtue of utilizing crop residues for rearing the animals.

In this paper different voluminous data like of stock market, rice production and milk production are collected in order to forecast the future values for the same using different ANN architectures and also to suggest the most efficient architecture among the implemented one.

II. RELATED WORK

Niall O'Connor, Michael G. Madden et al. [5] in their paper evaluated the effectiveness of using the external indicators, like commodity prices and currency exchange rates, in predicting the Dow Jones Industrial Average index. They also evaluated the performance of each technique using different domain-specific metrics.

Wei Shen , Xiaopen Guo , Chao Wu, Desheng Wu et al. [6] in their work have selected a radial basis function neural network (RBFNN) to train data and forecast the stock indices of the Shanghai Stock Exchange.

Manna Majumder, MD Anwar Hussian et al. [7] in their research work presented a computational approach for predicting the S&P CNX Nifty 50 Index. A neural network based model was used in predicting the direction of the movement of the closing value of the index. The model presented also confirmed that it could be used to predict price index value of the stock market. After studying the various features of the network model, they proposed an optimal model for the purpose of forecasting. The proposed model has used the preprocessed data set of closing value of S&P CNX Nifty 50 Index.

Milad Jasemi, Ali M. Kimiagari, A. Memariani et al. [8] in their paper have presented a new model to do stock market timing which used supervised feed forward neural network and the technical analysis of Japanese Candlestick.

Seyed Taghi Akhavan Niaki and Saeid Hoseinzade et al. [9] in their paper forecasted the daily direction of Standard & Poor's 500 (S&P 500) index using an artificial neural network (ANN).

Ratnadip Adhikari and R.K. Agrawal et al. [10] in their paper explored the outstanding ability of ANNs in recognizing & forecasting strong seasonal patterns without removing them from raw the data.

N. N. Jambhulkar et al. [3] in his research paper has described an Auto Regressive Integrated Moving Average (ARIMA) methodology and forecasted the rice production in the past study. In the present study, ARIMA stochastic modeling is used for describing rice production in Punjab. The yearly rice production data of Punjab from 1960-61 to 1999-2000 has been taken for model building and the data from 2000-01 to 2009-10 has been used for validation of the model. The best model has been selected based on the minimum Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values. It has been found that ARIMA (1, 1, 2) model described the rice production data in Punjab.

Dr. Sachin Kumar and Narendra Kumar et al. [4] in their paper have provided some modified techniques for time series forecasting for forecasting the yield of any crop year. They have used historical data of rice production for forecasting and validation. There were some limitations of the proposed techniques as Statistical methods show promising results when data is following some pattern. But due to some sudden hikes and falls in data the result is varying beyond threshold.

Satya Pal Ramasubramanian V. and S.C. Mehta et al. [11] in their paper made an attempt to forecast Milk production using Double Exponential Smoothing and ARIMA model. On validation of the forecasts from these two models, ARIMA performed better than the first one.

III. DATA MINING

Data mining is the process of handling information from databases, data warehouses, data marts, web, etc. which is hidden and cannot be seen directly. Data mining is used for a various purposes in private sectors. Data analysis techniques which have been traditionally used for such tasks include regression analysis, cluster analysis, numerical taxonomy, multidimensional analysis, other multivariate statistical methods, stochastic models, time series analysis and nonlinear estimation techniques [12].

Data mining is a process used by companies to turn raw data into useful and meaningful information. Data mining depends on effective data collection and warehousing as well as computer processing [13].

The main purpose of data mining is to develop decision making models for estimating the behaviors of the future values based on the analysis of the past activities [14].

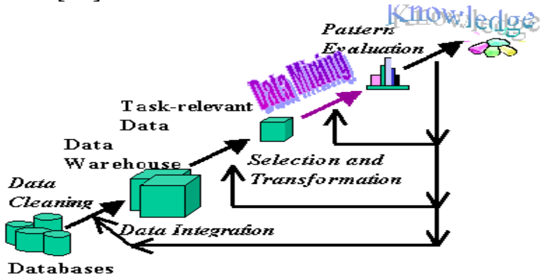

Fig 1: Data Mining Process [15]

The Knowledge Discovery in Databases process comprises of a few steps leading from raw data collections to some form of new knowledge. The iterative process consists of the following steps:

- **Data cleaning**: it also known as data cleansing. It is a phase in which missing data and noisy data are removed from the collection.
- **Data integration**: multiple data sources, like databases, data marts etc., may be combined in a common source.
- **Data selection**: the data is analyzed for its relevance and relevant data is retrieved from the data collection.
- **Data transformation**: also known as data consolidation, in this phase the selected data is transformed into normalized form which is appropriate for data mining.
- **Data mining**: it is the most important step in which different techniques are applied to extract patterns that are useful.
- **Pattern evaluation**: interesting patterns which are useful in representing knowledge are identified based on given measures.
- **Knowledge representation**: in this phase the discovered knowledge is visually represented to the user. It uses visualization techniques to help end users to understand and interpret the data mining results.

IV. NEURAL NETWORK AND ITS USE IN FORECASTING

Human brain is made up of a vast network of neurons, which are interconnected. The neuron is the basic and fundamental unit of the nervous system.

The **biological neuron** is a simple processing unit that receives and combines signals from other neurons through Dendrites.

The figure below illustrates the structure of a biological neuron.

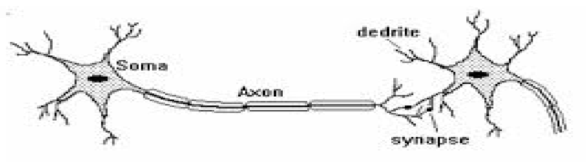

Fig2: A Biological Neuron [16]

The figure below shows the 3D view of a biological neural network.

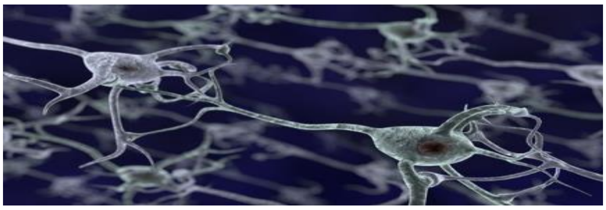

Fig 3: Biological Neural Network 3D [17]

An **artificial neuron** is an information processing unit that is fundamental to the operation of a neural network. The figure below shows an artificial neuron.

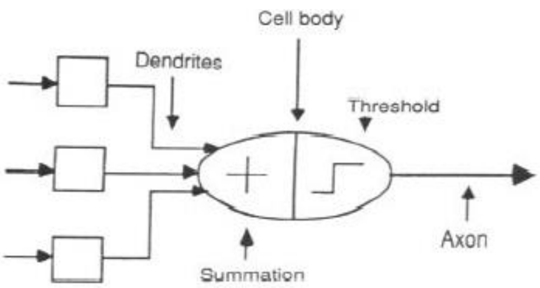

Fig 4: An Artificial Neuron [18]

An artificial neural network is composed of a number of interconnected units. Each unit has an input/output characteristic and an activation function. An ANN can be defined as a data- processing system which consists of a large number of simple, highly interconnected artificial neurons. Artificial neuron is also known as processing element in terms of ANN. These processing elements are organized into a sequence of layers i.e. input layer, hidden layer and output layer. An activation function is also required in ANN. The figure below shows the architecture of an artificial neural network.

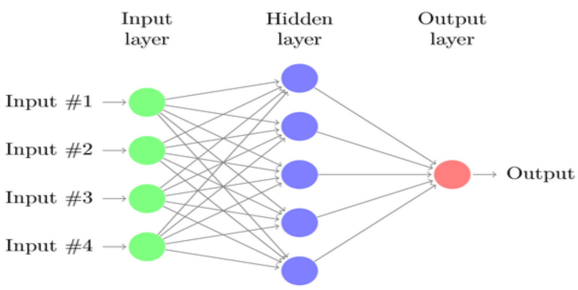

Fig 5: Architecture of Artificial Neural Network [19]

The four important features that are associated with artificial neural networks are:

- They learn by example,
- They constitute a distributed, associative memory,
- They are fault-tolerant, and
- They are capable of pattern recognition [20].

Artificial neural network have powerful pattern classification and pattern recognition capabilities they are able to learn from and generalise from experiences. ANNs have to be trained for its proper functioning i.e. a learning algorithm and the training set is required to train the artificial neural network so that it can predict the future value on the basis of training set.

Learning algorithms are broadly classified as:

- Supervised learning
- Unsupervised learning
- Reinforcement learning.

Now a day, ANN is considered to be one of the best tools to predict and forecast the future values of any field. ANN is the tool which is also used for forecasting the future values for voluminous data as it gives the effective and efficient results with less error values in comparison to other forecasting techniques. So in our study we too have used the ANN to predict the future values for voluminous data of S&P500, NASDAQ, DOW JONES, Rice & Milk production.

V. TIME SERIES

A time series is a sequence of data points, measured typically at successive points in time spaced at uniform time intervals. In other words, a time series is simply a sequence of numbers collected at regular intervals over a period of time, i.e. daily, weekly, monthly or annually . Time series are used in statistics, signal processing, pattern recognition, econometrics, mathematical finance, weather forecasting, earthquake prediction, control engineering, astronomy, and communications engineering.

Most of the generic time series definitions follow [21] Time series, series for short, is a sequence of numerical values indexed by increasing time units, e.g. a price of a commodity, such as oranges in a particular shop, indexed by the time when the price is checked. In the sequel, series' st return values refer to rt = log(st+T) − log(st), the return period T assumed 1, if not specified. Remarks about series distribution refer to the distribution of the returns series rt. A predictor forecasts a future values 0 t+T , having access only to past values si, i _ t, of this and usually other series. For the prediction to be of any value it has to be better than random.

Time series *forecasting* is the use of a model to predict future values based on previously observed values. Time series data are of two kinds:

- Continuous time series data
- Discrete time series data.

VI. RESEARCH METHODOLOGY

Artificial Neural Network is used to solve complex and large problems that people are facing in day to day life. Now-a-days Artificial Neural Networks are used as a tool to extract relevant information from the Data Warehouses which contain a large amount of data.

Algorithm:

I. Select the input variables as performance of Neural Network depends upon the selected input variables
II. Collect the data for n years
III. Pre-process the collected n years data for noisy data or for missing value data
IV. Design the ANN- decide the number of layers in ANN, number of neurons in the particular layer- We have taken different no. of input units, hidden layers, hidden units and output units each time.
V. Select the appropriate training algorithm for training the ANN- We have taken the Resilient algorithm to train the network. The historical data of stock market such as S&P500, NASDAQ, DOW JONES and that of Rice & Milk production in India has been used for training the ANN.
VI. Define the Transfer Function for hidden layer and output layer- We have used TANH activation function for hidden layers & output layer.
VII. Calculate the performance measures of the proposed ANN- We are calculating the Root Mean Square Error. RMS Error is the error between the actual value & the predicted value.

VII. PROPOSED WORK

One of the problems with the Back Propagation training algorithm is the degree to which the weights are changed. The goal of any training algorithm is to minimize the error function. The Resilient Back Propagation training algorithms (RPROP) only considers the sign of the gradient and not the value itself. Once the magnitude is discarded, this means it is only important if the gradient is positive, negative or near zero. For each weight, if there was a sign change of the partial derivative of the total error function compared to the last iteration, the update value for that weight is multiplied by a factor η^{-}, where $\eta^{-} < 1$. If the last iteration produced the same sign, the update value is multiplied by a factor of η^{+}, where $\eta^{+} > 1$. The update values are calculated for each weight in the above manner, and finally each weight is changed by its own update value, in the opposite direction of that weight's partial derivative, so as to minimise the total error function. η^{+} is empirically set to 1.2 and η^{-} to 0.5.

The Resilient Propagation training algorithm is usually the most efficient training algorithm provided by Encog framework used in this application for supervised feed-forward neural networks. One particular advantage to the RPROP algorithm is that it requires no setting of parameters before using it. There are no learning rates, momentum values or update constants that need to be determined. This is good because it can be difficult to determine the exact learning rate that might be optimal. Resilient back-propagation is considered the best algorithm, measured in terms of convergence speed, accuracy and robustness with respect to training parameters.

RPROP is one of the fastest weight update mechanisms.

VIII. EVALUATION AND SIMULATION PARAMETERS

We have developed a Feed forward Neural Network with different number of input units, hidden layers, number of nodes in each hidden layer and an output layer. Activation function –TANH for both hidden layer and output layer.

The different ANN architectures implemented in the study have the following combinations:

- Input layer, 1hidden layer 10 units and the output layer
- Input layer, 2 hidden layers, 20 hidden units and the output layer
- Input layer, 2 hidden layers, 30 hidden units and the output layer
- Input layer, 2 hidden layers, 41 hidden units and the output layer

These architectures are implemented using Encog Framework with support of Visual Studio. Different historical data has been used to forecast the future values of the same in order to help the investors in stock market, the economists by predicting the yield of the crops in the coming future or the milk production in the future. Data has been collected from different sources.

IX. RESULTS FOR DIFFERENT DATA SET PREDICTION

We have implemented and used different ANN architectures to ascertain parameters to predict precise data using different data sets. The historical stock market data has been taken from www.yahoofinance.com [22][23][24], Rice data has been collected from All India Rice Exporter Association www.airea.net [25] Milk production historical data has been collected from National Dairy Development Board www.nddb.org [26].

We have implemented different ANN architectures with different number of hidden layers and hidden units using Encog Framework and we have also calculated the RMS Error values for the same. Through this we have depicted a better ANN architecture to predict the voluminous data of different fields.

Table 1 Actual S&P500 and the predicted S&P500 with different ANN architecture

Date	Actual S&P500	ANN (1-10)	ANN (2-10)	ANN(2-20)	ANN(2-41)
1/4/2010	1115.1	1154.26	1124.21	1103.86	1083.86
1/5/2010	1132.99	1154.77	1121.67	1105	1088.21
1/6/2010	1136.52	1158.69	1129.16	1107.62	1090.61
1/7/2010	1137.14	1161.9	1129.85	1112.58	1089.33
1/8/2010	1141.69	1161.07	1130.64	1115.76	1090.82
1/11/2010	1144.98	1162.46	1132.58	1117.89	1092.27
1/12/2010	1146.98	1162.83	1133.64	1120.23	1092.88
1/13/2010	1136.22	1164.02	1137.04	1120.4	1091.13
1/14/2010	1145.68	1163.26	1136.44	1119.51	1091.97
1/15/2010	1148.46	1166.18	1135.15	1118.16	1093.81
1/19/2010	1136.03	1167.58	1137.07	1117.19	1094.97
1/20/2010	1150.23	1166.48	1138.07	1121.79	1094.26
1/21/2010	1138.04	1169.36	1141.36	1120.24	1095.8
1/22/2010	1116.48	1168.71	1137.85	1121.6	1095.71
1/25/2010	1091.76	1164.39	1137.96	1122.33	1097.25
1/26/2010	1096.78	1161.74	1136.84	1122.01	1099.52
1/27/2010	1092.17	1160.23	1136.28	1118.29	1099.27
1/28/2010	1097.5	1158.56	1132.93	1115.24	1094.79
1/29/2010	1084.53	1155.12	1132.5	1114.29	1091.83

The graph below shows the graphical representation of the actual & predicted S&P500 by different ANN architectures. This graph depicts that the ANN architecture with an input layer, 2 hidden layer, 10 hidden units and an output layer gives the most accurate predicted results. But until the RMS Error value is calculated we cannot say which architecture is better.

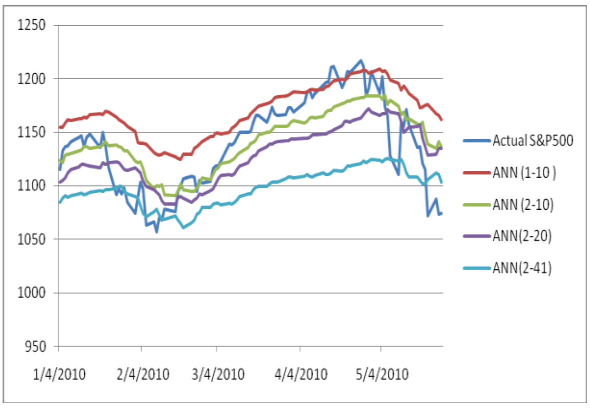

Fig 6: Result of actual & Predicted S&P500 by different ANN architecture

To calculate the RMS Error the following equation has been used

$$\text{RMS Errors} = \sqrt{\sum_{i=1}^{n}(actual - predicted)^2 / n}$$

Here is the graph which shows the comparison of the RMS Error values corresponding to different architectures of ANN and through this we can predict which ANN architecture is better to forecast S&P500.

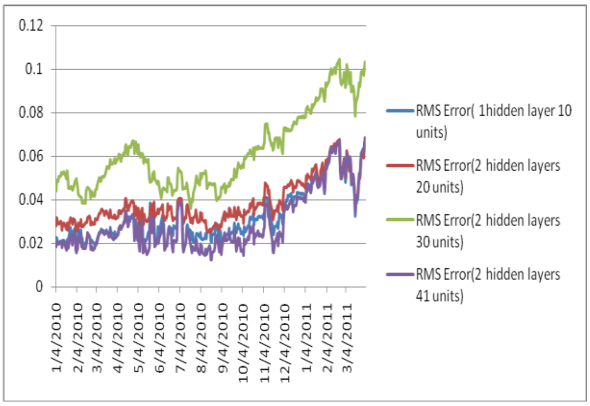

Fig 7: RMS Error values for different architecture

By analyzing the above graph we depict that the RMS Error values of ANN architecture with 2 hidden layers and 41 hidden units are less as compared to other Ann architectures. So for predicting S&P500 ANN architecture with 2 hidden layers and 41 hidden units is better than the other architectures.

Table 2 Actual NASDAQ and the predicted NASDAQ with different ANN architecture

Date	Actual Nasdaq	ANN (1-10)	ANN (2-10)	ANN(2-20)	ANN(2-41)
1/4/2010	2269.15	2092.2	2017.66	2359.12	2193.1
1/5/2010	2308.42	2110.13	2010.86	2365.15	2194.51
1/6/2010	2308.71	2122.93	2030.39	2375.58	2216.84
1/7/2010	2301.09	2134.13	2034.21	2399.09	2217.01
1/8/2010	2300.05	2132.43	2037.79	2411.58	2225.28
1/11/2010	2317.17	2134.71	2043.52	2418.56	2235.22
1/12/2010	2312.41	2128.98	2047.24	2426.59	2235.17
1/13/2010	2282.31	2124.51	2055.83	2427.37	2236.8
1/14/2010	2307.9	2122.88	2053.92	2428.16	2242.55
1/15/2010	2316.74	2138.34	2050.33	2418.07	2250.77
1/19/2010	2287.99	2143.72	2055.67	2417.61	2248.41
1/20/2010	2320.4	2135.96	2059.55	2436.66	2249.96
1/21/2010	2291.25	2145.37	2067.53	2428.69	2268.4
1/22/2010	2265.7	2147.35	2059.26	2434.77	2253.17
1/25/2010	2205.29	2139.82	2058.56	2436.62	2250.13
1/26/2010	2210.8	2142.88	2052.39	2437.15	2256.3
1/27/2010	2203.73	2155.33	2048.4	2416.92	2244.24
1/28/2010	2221.41	2160.78	2038.51	2408.52	2223.09
1/29/2010	2179	2152.98	2038.28	2409.29	2219.57

This graph shows the graphical representation of the actual & predicted NASDAQ by different ANN architectures. This graph depicts that the ANN architecture with an input layer, 2 hidden layer, 41 hidden units and an output layer gives the most accurate predicted results.

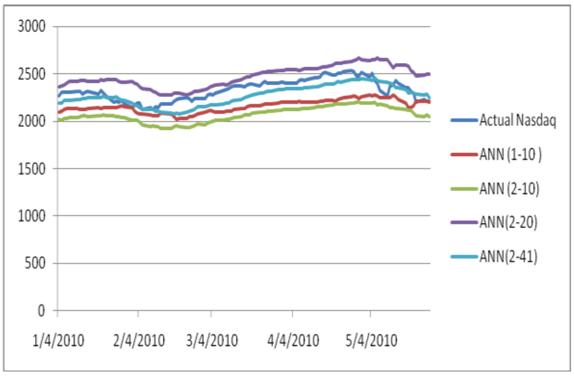

Fig 8: Result of actual & Predicted NASDAQ by different ANN architecture

The RMS Error value for the above ANN architectures is shown in the graph below

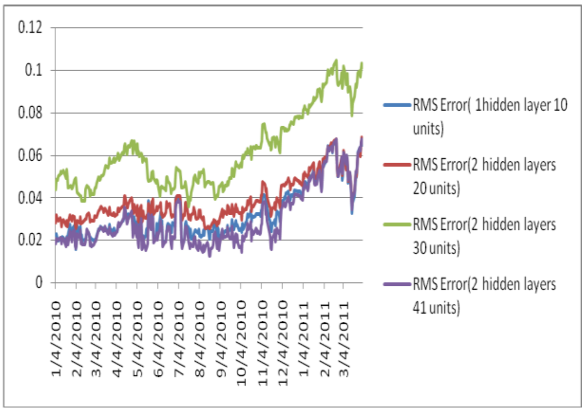

Fig 9: RMS Error values for different architecture

In the above graph, the ANN architecture with 2 hidden layers and 41 hidden units gives the minimum RMS Error values in comparison to others.

Table 3 Actual DOW JONES and the predicted DOW JONES with different ANN architecture

Date	Actual Dow Jones	ANN (1-10	ANN (2-10)	ANN(2-20)	ANN(2-41)
1/4/2010	10428.05	10410.55	10622.83	9494.09	10700.66
1/5/2010	10583.96	10406.08	10609.68	9489.92	10699.3
1/6/2010	10572.02	10436.78	10656.11	9486.19	10736.57
1/7/2010	10573.68	10462.04	10672.1	9508.45	10741.11
1/8/2010	10606.86	10457.06	10686.52	9499.06	10756.9
1/11/2010	10618.19	10469.05	10699.29	9491.29	10772.46
1/12/2010	10663.99	10472.84	10711.24	9499.09	10774.14
1/13/2010	10627.26	10490.59	10730.32	9509.12	10781.61
1/14/2010	10680.77	10482.97	10724.83	9529.56	10794.26
1/15/2010	10710.55	10497.72	10719.09	9528.12	10809.59
1/19/2010	10609.65	10511.32	10733.43	9538.85	10799.23
1/20/2010	10725.43	10507.09	10747.71	9534.29	10810.47
1/21/2010	10603.15	10527.71	10761.17	9532.67	10837.66
1/22/2010	10389.88	10519.68	10747.01	9538.37	10813.04
1/25/2010	10172.98	10488.38	10744	9506.61	10809.12
1/26/2010	10196.86	10462.06	10719.39	9477.35	10815.14
1/27/2010	10194.29	10434.83	10706.35	9456.36	10793.87
1/28/2010	10236.16	10424.5	10683.22	9446.62	10752.37
1/29/2010	10120.46	10397.05	10684.31	9426.25	10742.96

The graph below shows the graphical representation of the actual & predicted DOW JONES by different ANN architectures. The graph depicts that ANN with 2 hidden layers and 41 hidden units gives the more accurate predicted results.

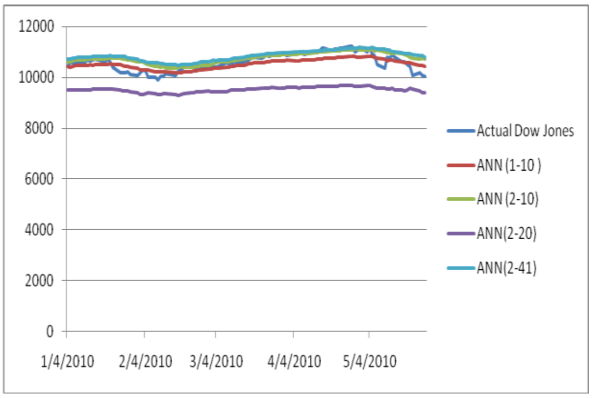

Fig 10: Result of actual & Predicted DOW JONES by different ANN architecture

The RMS Error value for the above ANN architectures is shown in the graph below

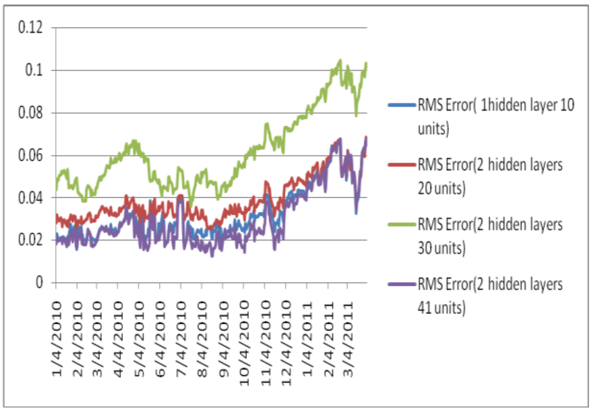

Fig 11: RMS Error values for different architecture

In the above graph, the ANN architecture with 2 hidden layers and 41 hidden units gives the minimum RMS Error values in comparison to others.

Table 4 Actual Milk Production and the predicted Milk Production with different ANN architecture

Year	Production (Million Tonnes)	ANN (1-10)	ANN (2-10)	ANN(2-20)	ANN(2-41)
1991-92	55.7	55.7	55.7	55.7	55.7
1992-93	58	55.7	55.7	55.7	55.7
1993-94	60.6	56.85	0	56.16	58
1994-95	63.8	58.1	59.965	57.048	60.6
1995-96	66.2	60.8	63.03	58.3984	63.8
1996-97	69.1	63.533333	65.56	64.63968	66.2
1997-98	72.1	66.366667	68.4	68.20794	69.1
1998-99	75.4	69.133333	71.355	71.32159	72.1
1999-2000	78.3	72.2	74.59	74.58432	75.4
2000-01	80.6	75.266667	77.555	77.55686	78.3
2001-02	84.4	78.1	79.995	79.99137	80.6
2002-03	86.2	81.1	83.525	83.51827	84.4
2003-04	88.1	83.733333	85.65	85.66365	86.2
2004-05	92.5	86.233333	87.63	87.61273	88.1
2005-06	97.1	88.933333	91.525	91.52255	92.5
2006-07	102.6	92.566667	95.96	95.98451	97.1
2007-08	107.9	97.4	101.27	101.2769	102.6
2008-09	112.2	102.53333	106.565	106.5754	107.9
2009-10	116.4	107.56667	111.075	111.0751	112.2
2011-12	127.9	116.8	120.51	120.507	121.8
2012-13	132.4	122.03333	126.41	126.4214	127.9

The graph below shows the graphical representation of the actual & predicted Milk Production in India by different ANN architectures. The graph depicts that ANN with 2 hidden layers and 41 hidden units gives the more accurate predicted results.

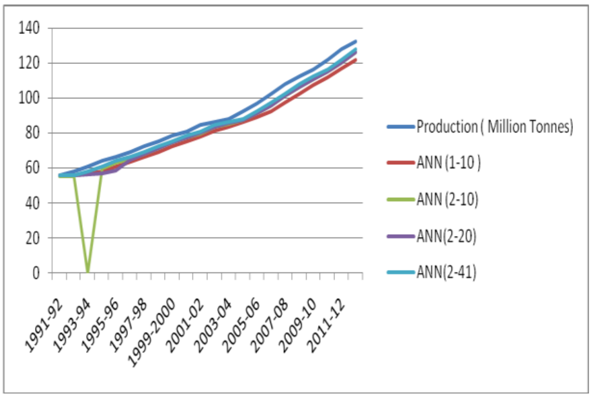

Fig 12: Result of actual and predicted Milk Production in India

The RMS Error value for the above actual and predicted data is given below. The graph shows the RMS Error values for all the different architectures and by analyzing this graph we can conclude that ANN architecture with 2 hidden layers and 41 hidden units gives minimum error values.

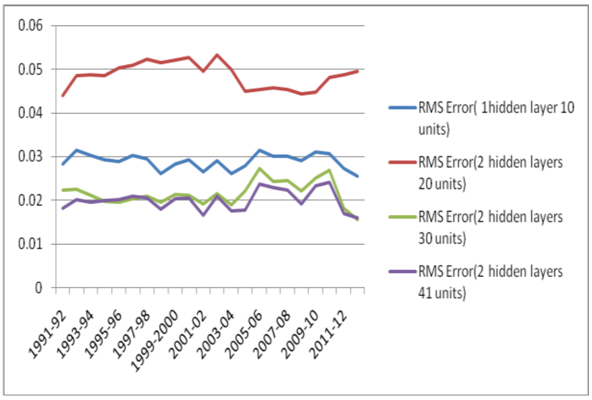

Fig 13: RMS Error values for different architecture

Table 5 Actual Rice Production and the predicted Rice Production with different ANN architecture

Year	Production (Million Tons)	ANN (1-10)	ANN (2-10)	ANN(2-20)	ANN(2-41)
1950-51	20.58	20.58	20.58	20.58	20.58
1951-52	21.3	20.58	20.58	20.58	20.58
1952-53	22.9	20.94	0	20.724	21.3
1953-54	28.21	21.5933333	22.544	21.1592	22.9
1954-55	25.22	24.1366667	27.068	22.56936	28.21
1955-56	27.56	25.4433333	25.5525	24.689872	25.22
1956-57	29.04	26.9966667	27.2415	26.985974	27.56
1957-58	25.53	27.2733333	28.627	28.629195	29.04
1958-59	30.85	27.3766667	26.158	27.329704	25.53
1959-60	31.68	28.4733333	29.9615	29.31275	30.85
1960-61	34.58	29.3533333	31.248	30.794023	31.68
1961-62	35.66	32.37	33.9585	33.426691	34.58
1962-63	33.21	33.9733333	35.299	35.046961	35.66
1963-64	37	34.4833333	33.646	33.952699	33.21
1964-65	39.31	35.29	36.3645	35.904649	37
1965-66	30.59	36.5066667	38.6585	38.220975	39.31
1966-67	30.44	35.6333333	32.2185	35.571608	30.59
1967-68	37.61	33.4466667	30.906	34.545286	30.44
1968-69	39.76	32.88	36.1835	35.286908	37.61

The graph below shows the graphical representation of the actual & predicted Rice Production in India by different ANN architecture.

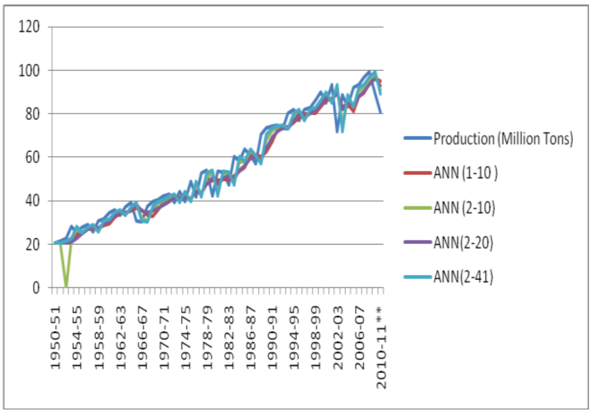

Fig 14 Result of actual and predicted Rice Production in India.

By analyzing the graph we can say that the Ann architecture with 2 hidden layers and 41 hidden units gives the accurate results.

The RMS Error value graph for different ANN architecture is shown below. The graph shows that the architecture with 2 hidden layers and 41 hidden units gives the minimum error value as compared to other architectures.

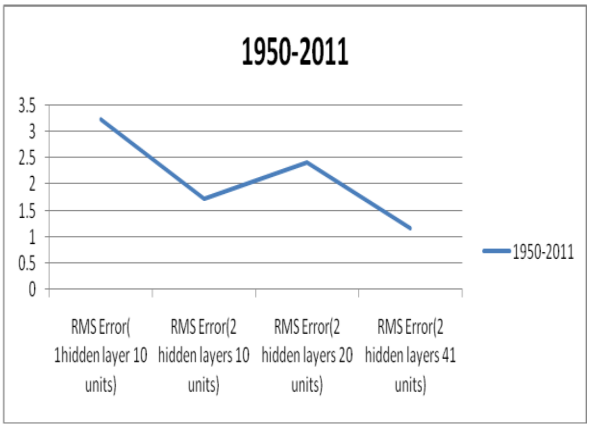

Fig 15 The RMS Error value

X. CONCLUSION

Artificial Neural Network is the need of today as it is the best tool to predict the future value. It is used in every field today like medical, schools, industries, etc.

In this paper the data is gathered from www.yahoofinance.com [22][23][24] and various other sources like www.airea.net [25] and www.nddb.org [26].

Time Series forecasting methods are based on the analysis of historical data. It makes the assumption that past historical patterns in data can be used to forecast future data values. We know that time series present for different fields can be sufficiently large, small or even no time series data available. We have used Resilient algorithm for the feed forward artificial neural network which affects the results and the efficiency of the predicted results. On the basis of our work we have proposed a new technique using neural network and data mining which gives better results after training the network.

Through this analysis using different data sets we can conclude that when the bulk of data is present the Feed Forward Artificial Neural Network with 2 hidden layers and 41 hidden units gives the minimum RMS Error which is the prime parameter to judge the efficiency of our implemented model. Less the RMS Error value more is the efficiency of the model.

This study includes the concept of data mining and neural network. The Root Mean Square Error value is calculated between the actual data and the predicted data. Future work can be calculating the other Error matrices which may be more accurate. This concept of the predicting the future value for stock market data can be used for forecasting some other data values which may be useful for the organization or in public interest.

References

[1] http://www.businessdictionary.com/definition/stock-exchange.html
[2] http://en.wikipedia.org/wiki/Stock_market
[3] N. N. Jambhulkar "*Modeling of Rice Production in Punjab using ARIMA Model*" International Journal of Scientific Research Volume 2, Issue 8, August 2013 , ISSN No 2277 – 8179
[4] Dr. Sachin Kumar and Narendra Kumar "*A Novel Method for Rice Production Forecasting Using Fuzzy Time Series*" IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 6, No 2, November 2012
[5] Niall O'Connor, Michael G. Madden *"A neural network approach to predicting stock exchange movements using external factors"* Knowledge-Based Systems 19 (2006) 371–378
[6] Wei Shen , Xiaopen Guo , Chao Wu, Desheng Wu "*Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm"2010*
[7] Manna Majumder, MD Anwar Hussian *"Forecasting of Indian Stock Market Index Using Artificial Neural Network*"
[8] Milad Jasemi, Ali M. Kimiagari, A. Memariani *"A modern neural network model to do stock market timing on the basis of the ancient investment technique of Japanese Candlestick"2010*
[9] Seyed Taghi Akhavan Niaki and Saeid Hoseinzade "*Forecasting S&P 500 index using artificial neural networks and design of experiments"* Journal of Industrial Engineering International 2013, 9:1
[10] Ratnadip Adhikari and R.K. Agrawal "*Forecasting strong seasonal time series with artificial neural network"* Journal of Scientific & Industrial Research Vol. 71, October 2012, pp. 657-666
[11] Satya Pal Ramasubramanian V. and S.C. Mehta *"Statistical Models for Forecasting Milk Production in India"* J. Ind. Soc. Agril. Statist. 61(2), 2007 : 80-83
[12] Sai Sumathi, S. N. Sivanandam , *Introduction to Data Mining and its Applications*, Vol. 29, ISBN 3-540-34350-4. Verlag Berlin Heidelberg: Springer, 2006.
[13] http://www.investopedia.com/terms/d/datamining.asp
[14] Koyuncugil, A.Serhan, Özgülbaş Nermin, "Data Mining: Data Mining: Using and Applications in Medicine and Healthcare," *Journal of Information Technology*, Vol .2, No. 2, pp.21-32, 2009.
[15] http://webdocs.cs.ualberta.ca/~zaiane/courses/cmput690/notes/Chapter1/
[16] http://neuralfuzzy.blogspot.in/2008/05/biological-neuron-model-and-artificial.html
[17] http://ns.scienceart.com/image/?id=2971&pagename=Neural_Network_3D
[18] http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html
[19] http://www.texample.net/media/tikz/examples/PDF/neural-network.pdf
[20] "Artificial Intelligence and Intelligent Systems" N.P.Padhy Oxford Higher Education.
[21] A. Refenes and M. Azema-Baras, "Neural network application in financial asset management", Neural Comput. Applicat., Vol . 2, PP . 13-39, 1994.
[22] http://finance.yahoo.com/q/hp?s=^GSPC&a=04&b=1&c=2000&d=04&e=1&f=2014&g=d
[23] http://finance.yahoo.com/q/hp?s=^IXIC&a=04&b=1&c=2000&d=04&e=1&f=2014&g=d
[24] http://finance.yahoo.com/q/hp?s=^DJI&a=04&b=1&c=2000&d=04&e=1&f=2014&g=d&z=66&y=0
[25] http://www.airea.net/page/62/statistical-data/all-india-area-production-and-yield-of-rice
[26] http://www.nddb.org/English/Statistics/Pages/Milk-Production.aspx

