INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

ORIGINAL ARTICLE

Corresponding author.
 Inam Ali Supro

Department of Computer Science, Shah Abdul Latif University, Sindh, Khairpur, Pakistan
inamsupro@yahoo.com

Rice yield prediction and optimization using association rules and neural network methods to enhance agribusiness

Inam Ali Supro[1]*, Javed Ahmed Mahar[1], Shahid Ali Mahar[1]

1 Department of Computer Science, Shah Abdul Latif University, Sindh, Khairpur, Pakistan

Abstract

Objectives: This study aims to implement data analytics and machine learning approaches to rice data and to establish association rules on fixed attributes and their correlations for yield prediction of crops. **Methods:** The data of rice crop is collected from district Larkana as per defined parameters: area, production, yield, temperature, rainfall, humidity and wind speed. The pre-processing operations are applied on prepared dataset to execute data analytics and machine learning algorithms. The processed data are then input into an Apriori algorithm for generating association rules. Neural network model is used to perform optimization on resulted association rules. **Findings:** Minimum support and confidence values equal to 3 and 80 respectively were set using Apriori algorithm on prepared rice dataset and obtained 88 association rules. Among them, results of 28 associated rules predicted `High Yield Production'. Neural network model is experimented to optimize the predicted yield of district Larkana through which overall network performance of 97.8% is calculated. Previously, rice yield data of Larkana were not statistically and digitally predicted and investigated. **Application :** Group of effective and well-built association rules of yield prediction are core outcome of this study which will be helpful for researchers, farmers and government officials to improve the productivity of rice crop.

Keywords: Rice data sets; Association rules; Yield prediction; Data optimization; Neural network

1 Introduction

Crop yield prediction is helpful in food production and management. Its role in the country's economy is inescapable[1]. Pakistan is a well-known agriculture based country. The agriculture sector of Pakistan contributes approximately 23.4% in its economy. Majority of the people are living in rural areas and purely dependent on agricultural activities[2]. Among the crops, rice has great importance because Pakistan has 4th position across the world, among the Rice exporters. Province of Sindh is the main contributor with approximately 25% rice produce in various districts especially in Larkana.

But most of the farmers are practicing traditional crop production methods instead of advanced technology and smart farming. The rapid changes in weather condition, results in the low crop productivity so that essential measures should be considered for yield prediction.

The current challenges and threats faced by the farmers are uncertain due to rapid changing conditions of weather, water shortages, poor irrigation facilities, uncontrolled cost and accurate yield prediction[3]. Therefore, it is necessary for the farmers to be equipped with smart farming particularly crop yield prediction using statistical and computational techniques. A research study proves that the yield crop prediction would facilitate farmers to make precautionary actions to improve productivity. Early prediction is possible through collection of previous experience of farmers, weather conditions and other influencing factors, by storing them in a large database with variety of parameters such as area, temperature, rainfall, humidity and wind speed.

Modern Technology is able to provide a lot of information on agricultural-related activities, through which necessary information can be extracted and sorted using data analyzing techniques. The big data analytics and other applications are used in this purpose to increase the agricultural yield production and reduce the expenses of the farming by taking appropriate decisions. Various machine learning techniques such as prediction, classification, regression and clustering are available for weather forecasting and crop yield prediction[4]. In this study, the rice data of district Larkana is collected from the Agriculture Statistics Department, Islamabad. These data are experimented to predict the yield using the Association Rule based approach. The Artificial Neural Network (ANN) Feed Forward Back propagation method of supervised learning are used to perform optimization on resulted association rules for decision-making.

1.1 Literature review

Data science is being utilized by most researchers in different areas of research including agriculture. Surya[5] carried a survey pertaining to the various factors of agriculture using the techniques of big data analytics and suggested that the implementation of big data analytics in agriculture gives better results. Furthermore, crop yield prediction is mainly preferred by the researchers using the Data Mining Techniques (DMTs). Rao[6] has used DMTs to envisage yield of crop in India. Many data mining algorithms are reported in literature to acquire the association rules. Some famous algorithms are AprioriTid, Apriori and Eclat implemented on different variables to find the association. The DMTs are also used by Ramakrishna[7] for agriculture land soil and reported the improvement in crop productivity. Specifically, an Apriori association rules were used to generate the advisory reports that facilitate decisions that support crop rotation, fertilizer requirements and harvesting procedures to the farmers. Supriya[8] used DMTs to analyze soil datasets for prediction of crops and developed complete system using K-Nearest Neighbor and Naive Bayes approaches for predicting the crop yield. Additionally, Zingade[9] presented an android system that used data analytics techniques i.e. linear regression to predict the most profitable crop in the current weather and soil conditions. The temperature, rainfall, soil parameters and past year production are used as selected attributes. The classification, clustering and association rule DMTs are analyzed by Chouhan[10] for analyzing the crops production.

During the literature review it is found that ANN is commonly used by the researchers for yield prediction and optimization. The ANN and statistical techniques are compared by Paswan[11]. Pandey[12] made prediction of potato yield crop using Generalized Regression Neural Network (GRNN) and a Radial Basis Function Neural Network (RBFNN). The leaf area index, biomass and plant height were used as input data parameters, while the yield of potato set as output dataset to train and test the NN. Author concluded that GRNN is a better predictor than RBFNN on the basis of quick learning capability and lower spread constant parameters. Niedbala[13] used ANN with Multi-Layer Perceptron (MLP) and presented three neural models for prediction of winter rapeseed yield. The temperature, precipitation and information about mineral fertilization are used to make the prediction of yield. The forecasting quality of the models is verified by determination of forecast errors. Similarly, Mathieu[14] used NN classifier for forecasting extreme corn yield losses caused by weather extremes. The average temperature and standardized precipitation evapo-transpiration index are used as set predictors. Moreover, Khaki[15] used deep NN to predict yield, check yield, and yield deference of corn hybrids from genotype and environment data. The study concludes that environmental factors had a greater effect on the crop yield than genotype.

1.2 Research methodology

Some specific and predefined steps are always required to finish the chosen research project. Figure 1 depicted the core research methodology adopted to accomplish the task of rice yield prediction and optimization. Selection of appropriate variables is the first task of this research project in which yield is selected as dependent variable and area, temperature, rainfall, humidity, wind speed and rice production are considered as independent variables. The required data is particularly collected on the basis of mentioned independent and dependent variables. The datasets are separately prepared for preprocessing the needed operations and further analytical experiments.
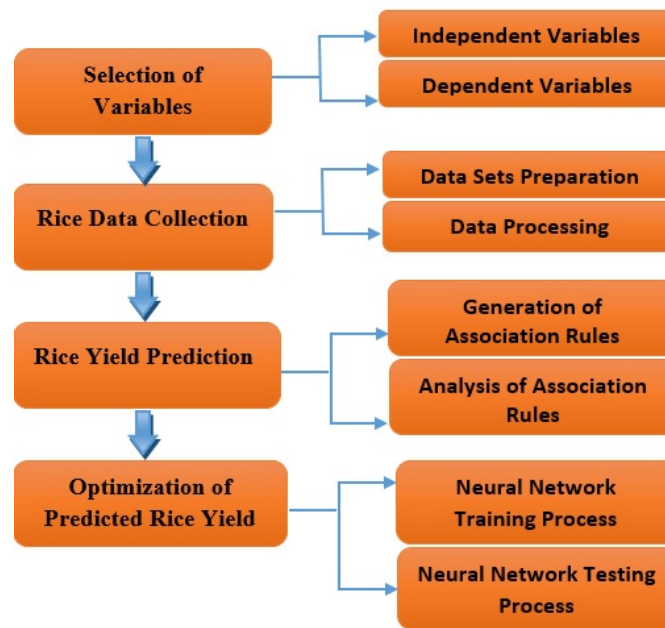
Fig 1. Research methodology

This research project is reported in two sections one deals with the yield prediction and other one is related with the optimization of predicted rice yield. The association rules based approach along with Apriori algorithm is used for rice yield prediction and various association rules are generated and analyzed using Frontline Excel Solver V2019. This software tool offers the facility of Analytic Solver through which association rules are automatically generated for further process. Furthermore, Matlab offers various optimization opportunities including the tool of ANN hence, NN model is proposed and implemented on the received yield prediction outcomes for optimization. It is better to develop own programming code with considerable parameters for prediction and obtain more accurate results but this process takes more time so that the built-in tool of Matlab is used in this research project.

1.3 Data collection and processing

The rice crop yield data (i.e. Area, Production and Yield) of Larkana District since year 1989 to 2015 collected from the Office of Director, Crop Reporting Services Centre, Hyderabad, Sindh and from the Sindh Government yearly published reports, namely Development Statistics of Sindh Report 2011, 2015 and 2017. The other important independent variable data (i.e. Temperature, Rainfall, Humidity and Wind speed) regarding Larkana District since year 1989 to 2018 from Office of Pakistan Meteorological Department, Karachi. Furthermore, the average data is received from the concerned department according to the selected variables. The process of rice crop is usually started from the month of June to November hence the data of these months are preferred for experiments. It is difficult to mention the total number of observations used in the yield prediction due to the average values of datasets. Hence, it is thoroughly stated that approximately 144 observations are experimented and optimized.

Many attributes are usually considered collecting the required data, out of which only the relevant attributes are selected and combined to develop rice data set of Larkana District for the laboratory experiments. The agricultural data and other relevant data variables are organized and combined. The agricultural data includes Area of Rice under Cultivation (in Hectares), Rice Production (in Million Tons) and Rice Yield (in KG's per Hectares). While, other important variables include Temperature (in Degree Celsius), Annual Rainfall (in Milli Meters), Humidity (in Percentage) and Wind speed (in Knots).

All manually collected and published data were transferred into spreadsheet/computerized file, so that rice dataset can be input to the machine for the development of association rules. The Rice is Kharif Crop, its sowing in Larkana District starts in the month of June and harvesting starts in the November month of every year. Hence, average value of maximum temperature, average monthly amount of rainfall, average monthly relative humidity and average monthly wind speed of June to November of every year is taken for further preprocessing. The prepared datasets from the year 2010 to 2015 are given in Table 1.

Machine language algorithms process the data, when it is not in continuous form of data. The discretization is the process that allows making the values in limited number of possible states. The output of discretization is the file containing the ordered

Table 1. Prepared rice yield dataset

Year	μ Area	μ Temperature	μ Rainfall	μ Humidity	μ Wind speed	μ Production	μ Yield
2010	97823	36.88	42.50	45.67	3.57	383600	3921.37
2011	98935	35.17	28.75	49.83	3.50	388447	3926.28
2012	93944	37.23	36.83	46.00	3.72	362100	3854.42
2013	98009	37.20	16.27	51.00	4.27	322418	3289.68
2014	102134	38.17	0.67	41.83	3.82	295284	2891.14
2015	105223	36.95	13.22	48.00	3.27	409470	3891.45

and discrete values. The developed dataset is based on several attributes having continuous data. In this stage, discretization is performed on continues data of all attributes of rice crop dataset of Larkana District. The attributes: area, temperature, rainfall, humidity, wind speed, production and yield are discretized into least, average, and utmost on particularly fixed threshold value with preferred attributes. The set threshold values for all variables are given in Table 2.

Table 2. Setting of threshold values

Attributes	High	Medium	Low
Area	Above 150,000	between 100,000 to 150,000	Below 100,000
Temperature	Above 36	between 20 to 36	Below 20
Rainfall	Above 66	between 33 to 66	Below 33
Humidity	Above 50	between 45 to 50	Below 45
Wind speed	Above 3.5	between 2.5 to 3.5	Below 2.5
Production	Above 450,000	between 250,000 to 450,000	Below 250,000
Yield	Above 3000	between 1500 to 3000	Below 1500

2 Association Rules

Association Rules is one of the incredibly important concepts of machine learning commonly used by the researchers to solve the problems pertaining with big data analytics including crop yield prediction. The association rules sufficiently help to uncover all the relationships between the items from huge generated databases. For this, Apriori algorithm is used for getting empirical results. Moreover, association rules are if-then statements that help to show the probability of relationships between data items within large data sets stored in databases. An association rule has two parts: an antecedent (if) and a consequent (then), both of which are a list of items. An antecedent is an item found within the data, while a consequent is an item found in combination with the antecedent. Various metrics are in place to help us understand the strength of association between antecedent and the consequent.

2.1 Experiments with Apriori algorithm

The searching of widespread patterns is vital and Apriori algorithm allows significantly doing the task specifically among the items which are already stored in the developed database. This algorithm is utilized effectively and association rules were successfully and digitally produced with chosen item sets. Hence, the pre-processed and discretized rice dataset of district Larkana is given as input to this preferred algorithm for getting the frequent item sets and strong association rules. The process of this algorithm is started by creating the basic item set with the association of single item with all available item sets. On the basis of the candidate set of item the most useful sets of items are created by shortening all the items whose support computations do not fully satisfy the least value of support threshold. Hence, the standard item set consequently acquired items that support values are continuously increasing. Furthermore, strong association rules obtained from the experiments are examined to comprehend the correlation surrounded by the variables i.e. area, temperature, rainfall, humidity, wind speed, production and yield.

The established list of 88 various association rules auto generated by using Apriori algorithm on Larkana Rice Dataset. The auto generated list of association rules is divided into two parts; Tables 3 and 4 presented 1 to 44 and 45 to 88 rule lists respectively. Both tables contain Rule id, A-support, C-support, Support, Confidence, Lift-ratio, Antecedent and Consequent

attributes. The attribute Rule id identifies the rules in the table and Support, Confidence, Lift-ratio, Antecedent and Consequent are core parameters, while values of A-support, C-support parameters is not the scope of this research so that is not intensely discussed.

2.2 Measurements of support and confidence levels

The results are obtained using an Apriori algorithm contains the list of association rules. It is tried to achieve maximum possible combinations and stronger association rules among the items in the rice dataset by setting the minimum support and confidence values. The item sets are denoted with the variable X and the support is represented as supp(X). The X is always described as the proportion of the all possible transactions across the all data sets that hold the calculated item set. After computing the various support levels, it is a correct time to calculate the level of confidence using Eq.1 given below.

$$conf\left(X \rightarrow Y\right) = supp(X \cup Y)/supp(X) \tag{1}$$

It is noted here that the computational process of supp(X∪Y) actually represents the support for the occurrences of the both transactions X and Y. Furthermore, confidence is usually interpreted as the probability estimation P (Y|X) through which left and right sides of the obtained rules are calculated. Then, the left side of the rule is calculated using the Eq.2 given below to observe support percentage of the variable X and Y.

$$lift\left(X \rightarrow Y\right) = \frac{supp(X \cup Y)}{supp\left(X\right) \ X \ supp(Y)} \tag{2}$$

Only two types of support values were found while the experimenting. Figure 2 presented the support values adjacent to the rules. The highest support value is 4 and the least support value is 3. Surprisingly, rule numbers 3, 9, 11, 27, 28 and 30 have the highest support value which is 4 and remaining rules holds the least support value, hence the minimum support value is set with 3.
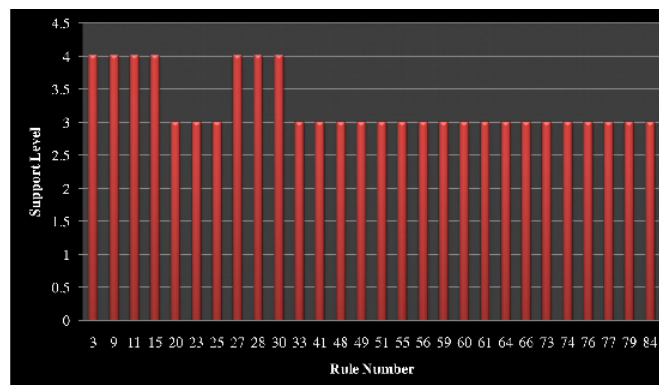

Fig 2. Calculated support levels

Similar to the calculation of Support, the confidence value is calculated to observe the relation between the rules as well as fixed variable. Figure 3 presents the calculated confidence values with all the generated rules. The least confidence value of 80 is observed with the rules 9, 15 and 28. All the other rules represent the maximum confidence value which is 100. Keeping the facts, minimum confidence value equals to 80 is fixed for the analysis.

2.3 Predicted rice yield

Tables 3 and 4 presented the outcomes of the association rules specifically in terms of the support and confidence level of the rules. In many places, the value of temperature, wind speed and other variables are very high as compared to variable yield. Hence, the purified list of those rules having yield at high level is separated for realizing the impact of other variables on the dependent variable yield.

The generated list of 28 strong filtered association rules that predict the Rice Yield at maximum level, with their Antecedent, Consequent, Support and Confidence values is given in Table 5. The value of support attribute shows the participation of item set given in Antecedent of each rule. The value of confidence expresses the support for occurrences of transactions where both

Table 3. Auto generated association rules from 1 to 44

Rule ID	A-Support	C-Support	Support	Confidence	Lift-Ratio	Antecedent	Consequent
Rule 1	4	5	4	100	1.2	[A-Low]	[P-Medium]
Rule 2	5	4	4	80	1.2	[P-Medium]	[A-Low]
Rule 3	4	5	4	100	1.2	[A-Low]	[Y-High]
Rule 4	5	4	4	80	1.2	[Y-High]	[A-Low]
Rule 5	5	4	4	80	1.2	[T-High]	[W-High]
Rule 6	4	5	4	100	1.2	[W-High]	[T-High]
Rule 7	5	5	4	80	0.96	[T-High]	[P-Medium]
Rule 8	5	5	4	80	0.96	[P-Medium]	[T-High]
Rule 9	5	5	4	80	0.96	[T-High]	[Y-High]
Rule 10	5	5	4	80	0.96	[Y-High]	[T-High]
Rule 11	4	5	4	100	1.2	[H-Medium]	[Y-High]
Rule 12	5	4	4	80	1.2	[Y-High]	[H-Medium]
Rule 13	4	5	4	100	1.2	[W-High]	[P-Medium]
Rule 14	5	4	4	80	1.2	[P-Medium]	[W-High]
Rule 15	5	5	4	80	0.96	[P-Medium]	[Y-High]
Rule 16	5	5	4	80	0.96	[Y-High]	[P-Medium]
Rule 17	3	4	3	100	1.5	[A-Low,T-High]	[W-High]
Rule 18	3	5	3	100	1.2	[A-Low,W-High]	[T-High]
Rule 19	3	5	3	100	1.2	[A-Low,T-High]	[P-Medium]
Rule 20	3	5	3	100	1.2	[A-Low,T-High]	[Y-High]
Rule 21	3	5	3	100	1.2	[A-Low,H-Medium]	[P-Medium]
Rule 22	3	4	3	100	1.5	[H-Medium, P-Medium]	[A-Low]
Rule 23	3	5	3	100	1.2	[A-Low,H-Medium]	[Y-High]
Rule 24	3	5	3	100	1.2	[A-Low,W-High]	[P-Medium]
Rule 25	3	5	3	100	1.2	[A-Low,W-High]	[Y-High]
Rule 26	3	4	3	100	1.5	[W-High,Y-High]	[A-Low]
Rule 27	4	4	4	100	1.5	[A-Low]	[P-Medium,Y-High]
Rule 28	5	4	4	80	1.2	[P-Medium]	[A-Low,Y-High]
Rule 29	5	4	4	80	1.2	[Y-High]	[A-Low,P-Medium]
Rule 30	4	5	4	100	1.2	[A-Low,P-Medium]	[Y-High]
Rule 31	4	5	4	100	1.2	[A-Low,Y-High]	[P-Medium]
Rule 32	4	4	4	100	1.5	[P-Medium,Y-High]	[A-Low]
Rule 33	3	5	3	100	1.2	[T-High,H-Medium]	[Y-High]
Rule 34	5	4	4	80	1.2	[T-High]	[W-High,P-Medium]
Rule 35	4	4	4	100	1.5	[W-High]	[T-High,P-Medium]
Rule 36	5	4	4	80	1.2	[P-Medium]	[T-High,W-High]
Rule 37	4	5	4	100	1.2	[T-High,W-High]	[P-Medium]
Rule 38	4	4	4	100	1.5	[T-High,P-Medium]	[W-High]
Rule 39	4	5	4	100	1.2	[W-High,P-Medium]	[T-High]
Rule 40	3	5	3	100	1.2	[W-High,Y-High]	[T-High]
Rule 41	3	5	3	100	1.2	[H-Medium,P-Medium]	[Y-High]
Rule 42	3	5	3	100	1.2	[W-High,Y-High]	[P-Medium]
Rule 43	3	4	3	100	1.5	[A-Low,T-High]	[W-High,P-Medium]
Rule 44	3	4	3	100	1.5	[A-Low,W-High]	[T-High,P-Medium]

Table 4. Auto generated association rules from 45 to 88

| Rule ID	A-Support	C-Support	Support	Confidence	Lift-Ratio	Antecedent	Consequent
Rule 45	3	5	3	100	1.2	[A-Low,T-High,W-High]	[P-Medium]
Rule 46	3	4	3	100	1.5	[A-Low,T-High,P-Medium]	[W-High]
Rule 47	3	5	3	100	1.2	[A-Low,W-High,P-Medium]	[T-High]
Rule 48	3	3	3	100	2	[A-Low,T-High]	[W-High,Y-High]
Rule 49	3	4	3	100	1.5	[A-Low,W-High]	[T-High,Y-High]
Rule 50	3	3	3	100	2	[W-High,Y-High]	[A-Low,T-High]
Rule 51	3	5	3	100	1.2	[A-Low,T-High,W-High]	[Y-High]
Rule 52	3	4	3	100	1.5	[A-Low,T-High,Y-High]	[W-High]
Rule 53	3	5	3	100	1.2	[A-Low,W-High,Y-High]	[T-High]
Rule 54	3	4	3	100	1.5	[T-High,W-High,Y-High]	[A-Low]
Rule 55	3	4	3	100	1.5	[A-Low,T-High]	[P-Medium,Y-High]
Rule 56	3	5	3	100	1.2	[A-Low,T-High,P-Medium]	[Y-High]
Rule 57	3	5	3	100	1.2	[A-Low,T-High,Y-High]	[P-Medium]
Rule 58	3	4	3	100	1.5	[T-High,P-Medium,Y-High]	[A-Low]
Rule 59	3	4	3	100	1.5	[A-Low,H-Medium]	[P-Medium,Y-High]
Rule 60	3	4	3	100	1.5	[H-Medium,P-Medium]	[A-Low,Y-High]
Rule 61	3	5	3	100	1.2	[A-Low,H-Medium,P-Medium]	[Y-High]
Rule 62	3	5	3	100	1.2	[A-Low,H-Medium,Y-High]	[P-Medium]
Rule 63	3	4	3	100	1.5	[H-Medium,P-Medium,Y-High]	[A-Low]
Rule 64	3	4	3	100	1.5	[A-Low,W-High]	[P-Medium,Y-High]
Rule 65	3	4	3	100	1.5	[W-High,Y-High]	[A-Low,P-Medium]
Rule 66	3	5	3	100	1.2	[A-Low,W-High,P-Medium]	[Y-High]
Rule 67	3	5	3	100	1.2	[A-Low,W-High,Y-High]	[P-Medium]
Rule 68	3	4	3	100	1.5	[W-High,P-Medium,Y-High]	[A-Low]
Rule 69	3	4	3	100	1.5	[W-High,Y-High]	[T-High,P-Medium]
Rule 70	3	5	3	100	1.2	[T-High,W-High,Y-High]	[P-Medium]
Rule 71	3	4	3	100	1.5	[T-High,P-Medium,Y-High]	[W-High]
Rule 72	3	5	3	100	1.2	[W-High,P-Medium,Y-High]	[T-High]
Rule 73	3	3	3	100	2	[A-Low,T-High]	[W-High,P-Medium,Y-High]
Rule 74	3	3	3	100	2	[A-Low,W-High]	[T-High,P-Medium,Y-High]
Rule 75	3	3	3	100	2	[W-High,Y-High]	[A-Low,T-High,P-Medium]
Rule 76	3	4	3	100	1.5	[A-Low,T-High,W-High]	[P-Medium,Y-High]
Rule 77	3	3	3	100	2	[A-Low,T-High,P-Medium]	[W-High,Y-High]
Rule 78	3	4	3	100	1.5	[A-Low,T-High,Y-High]	[W-High,P-Medium]
Rule 79	3	4	3	100	1.5	[A-Low,W-High,P-Medium]	[T-High,Y-High]
Rule 80	3	4	3	100	1.5	[A-Low,W-High,Y-High]	[T-High,P-Medium]
Rule 81	3	4	3	100	1.5	[T-High,W-High,Y-High]	[A-Low,P-Medium]
Rule 82	3	3	3	100	2	[T-High,P-Medium,Y-High]	[A-Low,W-High]
Rule 83	3	3	3	100	2	[W-High,P-Medium,Y-High]	[A-Low,T-High]
Rule 84	3	5	3	100	1.2	[A-Low,T-High,W-High,P-Medium]	[Y-High]
Rule 85	3	5	3	100	1.2	[A-Low,T-High,W-High,Y-High]	[P-Medium]
Rule 86	3	4	3	100	1.5	[A-Low,T-High,P-Medium,Y-High]	[W-High]
Rule 87	3	5	3	100	1.2	[A-Low,W-High,P-Medium,Y-High]	[T-High]
Rule 88	3	4	3	100	1.5	[T-High,W-High,P-Medium,Y-High]	[A-Low]
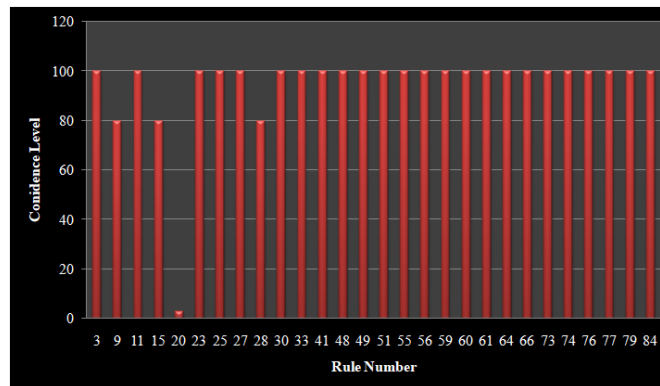
Fig 3. Calculated confidence levels

Antecedent and Consequent appears true. Basically, the value of confidence tells the probability of how many times the rule has become correct using the provided dataset. While the value of lift-ratio shows the ratio of observed support to the Antecedent and Consequent were found independent.

2.4 Neural network c lassification

Various machine learning approaches have been used for crop yield prediction as well as optimization. The NN approach has great importance and gives acceptable level accuracy hence, this approach is used for optimization of rice yield data. With the help of the neural network, the process of the data optimization is performed on the extracted information such as Area, Temperature, Rainfall and the Humidity against the Yield and Production which are the targeted parameters. Optimization performed through the neural network using the feed forward back propagation method uses the supervised learning. However, the application of neural network is used in another way by Gandhi [16] and Khaki [17] for crop yield prediction. It is considered that the algorithm is provided with samples of inputs and outputs. Based on the supervised learning, the network calculates the error. The major goal of this algorithm is that it is used to adjust the input weights in order to reduce the error rate so that process of learning and training data is achieved by NN. Figure 4 is the depiction of developed neural network for optimization of the predicted rice yield.
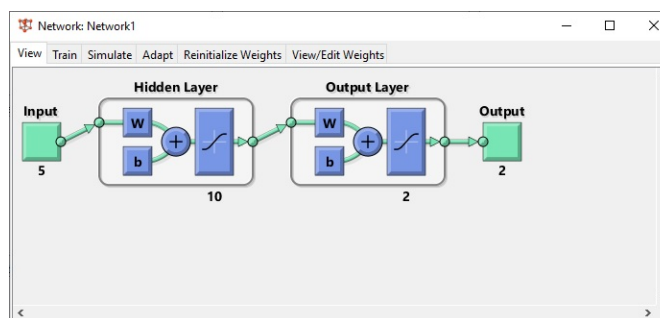

Fig 4. Proposed neural network

The developed network is based on the 5 input and 2 output values. Area, Temperature, Rainfall, Humidity and Wind Speed are the factor effecting yield and production performance. The remote sensing data is also used by Putri [18] with specific variables. Thus, these five variables are selected as input variables while Yield and Production are set as targeted values to train the network. Furthermore, ten hidden neurons are fixed as hyper parameters for developed network to get the accurate and reliable results. Some researchers argued that convolutional neural network is more appropriate than the back propagation network due to the time dependency [19]. In this project, time may be not considered in that way because it is the time period of the rice cropping so that the proposed neural network is sufficient enough for such kind of systems. The training and testing phase performed using a dataset made in excel work sheet later imported in Matlab programming language by creating 2 arrays of 5x6 and 2x6 as shown in Figure 5.

Table 5. Filtered association rules representing high predicted rice yields

Rule ID	Support	Confidence	Lift-Ratio	Antecedent	Consequent
Rule 3	4	100	1.2	[A-Low]	[Y-High]
Rule 9	4	80	0.96	[T-High]	[Y-High]
Rule 11	4	100	1.2	[H-Medium]	[Y-High]
Rule 15	4	80	0.96	[P-Medium]	[Y-High]
Rule 20	3	100	1.2	[A-Low,T-High]	[Y-High]
Rule 23	3	100	1.2	[A-Low,H-Medium]	[Y-High]
Rule 25	3	100	1.2	[A-Low,W-High]	[Y-High]
Rule 27	4	100	1.5	[A-Low]	[P-Medium,Y-High]
Rule 28	4	80	1.2	[P-Medium]	[A-Low,Y-High]
Rule 30	4	100	1.2	[A-Low,P-Medium]	[Y-High]
Rule 33	3	100	1.2	[T-High,H-Medium]	[Y-High]
Rule 41	3	100	1.2	[H-Medium,P-Medium]	[Y-High]
Rule 48	3	100	2	[A-Low,T-High]	[W-High,Y-High]
Rule 49	3	100	1.5	[A-Low,W-High]	[T-High,Y-High]
Rule 51	3	100	1.2	[A-Low,T-High,W-High]	[Y-High]
Rule 55	3	100	1.5	[A-Low,T-High]	[P-Medium,Y-High]
Rule 56	3	100	1.2	[A-Low,T-High,P-Medium]	[Y-High]
Rule 59	3	100	1.5	[A-Low,H-Medium]	[P-Medium,Y-High]
Rule 60	3	100	1.5	[H-Medium,P-Medium]	[A-Low,Y-High]
Rule 61	3	100	1.2	[A-Low,H-Medium,P-Medium]	[Y-High]
Rule 64	3	100	1.5	[A-Low,W-High]	[P-Medium,Y-High]
Rule 66	3	100	1.2	[A-Low,W-High,P-Medium]	[Y-High]
Rule 73	3	100	2	[A-Low,T-High]	[W-High,P-Medium,Y-High]
Rule 74	3	100	2	[A-Low,W-High]	[T-High,P-Medium,Y-High]
Rule 76	3	100	1.5	[A-Low,T-High,W-High]	[P-Medium,Y-High]
Rule 77	3	100	2	[A-Low,T-High,P-Medium]	[W-High,Y-High]
Rule 79	3	100	1.5	[A-Low,W-High,P-Medium]	[T-High,Y-High]
Rule 84	3	100	1.2	[A-Low,T-High,W-High,P-Medium]	[Y-High]

Using the above mentioned values the network developed on a method known as Feed Forward Back propagation Method. Data was optimized to generate the best output of Yield and Production. However, during the training process the best output is generated at the 7th epoch by the network out of 1000 which is shown in Figure 6. The batch training performed during which, samples pass through the learning algorithm before updating the weights at the learning rate of 0.001/epoch. While the Figure 7 showing the gradient ration generated by network which is 264971.27 up to the 7th epoch. The validation check is also shown in the same figure which is a total of 6 in 7 epochs.

The selected dataset is used for training as well testing of network. Experimented results showed that admirable outcomes were achieved during training and testing phase. In Figure 8 it is demonstrated that best validation performance results achieved by the network are 648 at the 1st epoch with respect to its Mean Square Error (MSE). Also, the graph depicted in the Figure 9 shows training as well testing performance against received MSE. While the overall network performance shows that 99.8% received during the network training and overall optimized results achieved are 97.8%.

3 Discussion

The already available rice data of Larkana district is experimented using Apriori algorithm by considering the parameters: area, temperature, rainfall, humidity, wind speed, production and yield data of six years. It is noted that the last two variables are
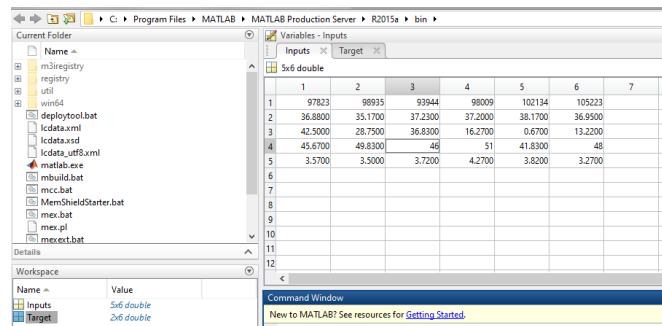
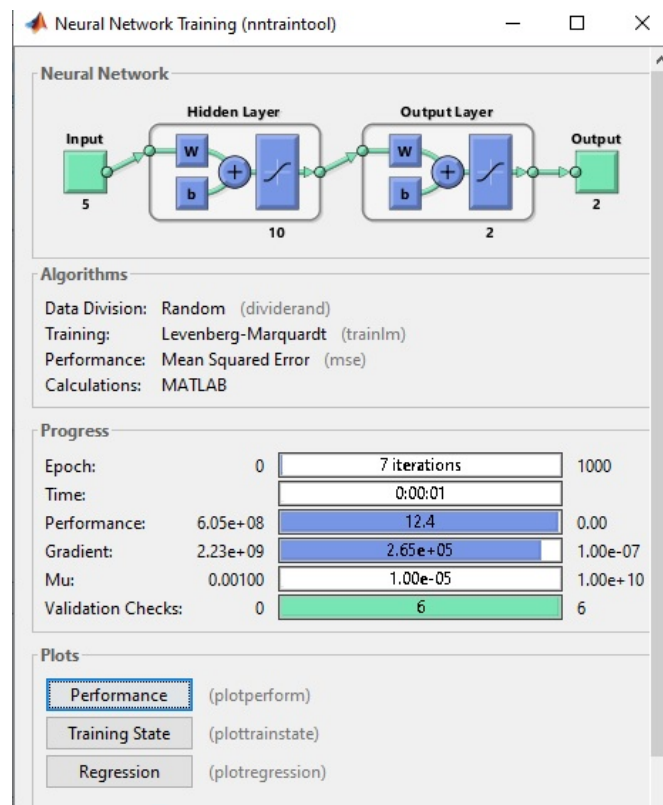**Fig 5.** Developed dataset for network

Fig 6. Execution process of developed network

fixed as targeted variables. The rules that predict the rice yield with satisfactory level of confidence are described in Table 6. After careful analysis of the rules and predicted rice yield, it is realized that many factors directly affect the yield of rice crop. For instance, rule number 3 defines that if the area is low then yield is high this means that the farmer is not able to provide the sufficient water to the rice crop. On the other hand, with the reflection of rule number 9, when the parameter temperature is high as compared to normal level then yield is also high because low temperature reduces the production of rice. The rule 11 provided the surprising results belonging with the humidity because the medium percentage of this factor gives maximum rice yield.

But, rule number 23 tells us that low area and medium level humidity is also a cause of high yield. The analysis is consciously realized with the rule number 25 in a way that if area is low and wind speed is high then yield is high. We have some parameters having directed relation such as area and production for example, rule number 30 defines if area is low and production is medium then yield is high. The rice yield is increased when the temperature is soaring and the humidity is average (see rule number 33). The rule 41 is characterized in a way that when both factors humidity and production are at appropriate levels
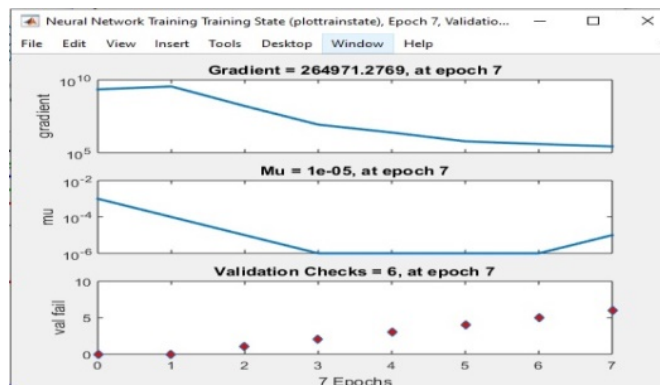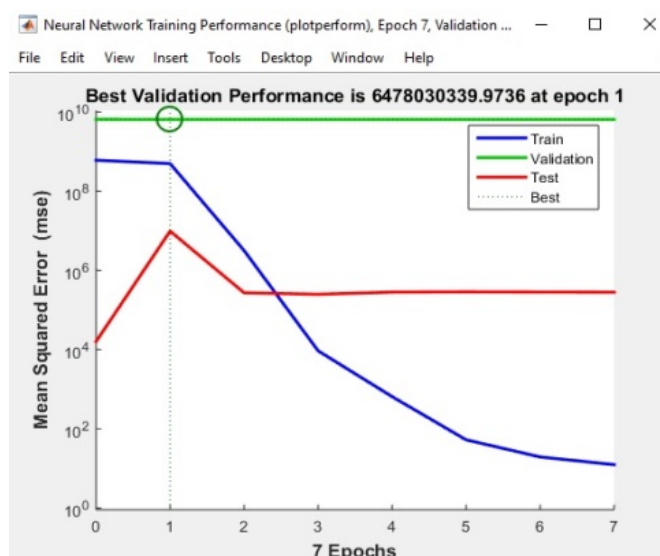
Fig 7. Gradient result
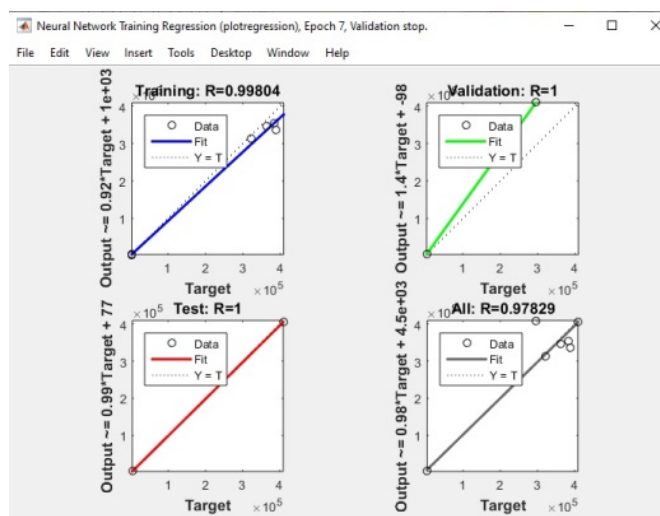

Fig 8. Validation performance

Fig 9. Overall network performances

Table 6. Rules predicted rice yield with satisfactory level

Rule ID	Confidence	Antecedent	Consequent
Rule 3	100	[A-Low]	[Y-High]
Rule 9	80	[T-High]	[Y-High]
Rule 11	100	[H-Medium]	[Y-High]
Rule 20	100	[A-Low,T-High]	[Y-High]
Rule 23	100	[A-Low,H-Medium]	[Y-High]
Rule 25	100	[A-Low,W-High]	[Y-High]
Rule 30	100	[A-Low,P-Medium]	[Y-High]
Rule 33	100	[T-High,H-Medium]	[Y-High]
Rule 41	100	[H-Medium,P-Medium]	[Y-High]
Rule 51	100	[A-Low,T-High,W-High]	[Y-High]
Rule 55	100	[A-Low,T-High]	[P-Medium,Y-High]
Rule 56	100	[A-Low,T-High,P-Medium]	[Y-High]
Rule 59	100	[A-Low,H-Medium]	[P-Medium,Y-High]
Rule 61	100	[A-Low,H-Medium,P-Medium]	[Y-High]
Rule 64	100	[A-Low,W-High]	[P-Medium,Y-High]
Rule 66	100	[A-Low,W-High,P-Medium]	[Y-High]
Rule 73	100	[A-Low,T-High]	[W-High,P-Medium,Y-High]
Rule 74	100	[A-Low,W-High]	[T-High,P-Medium,Y-High]
Rule 76	100	[A-Low,T-High,W-High]	[P-Medium,Y-High]
Rule 77	100	[A-Low,T-High,P-Medium]	[W-High,Y-High]
Rule 79	100	[A-Low,W-High,P-Medium]	[T-High,Y-High]
Rule 84	100	[A-Low,T-High,W-High,P-Medium]	[Y-High]

then yield is high. With the focusing of rule 51, it is found that if area, temperature and wind speed are touching the high level then yield is increase at acceptable level. The rule number 55 defines that if area is low and temperature is high then production is medium and yield is high but when the area is low and temperature is high and production is medium then yield is utmost showed by rule 56. It is also examined that using the rule number 59 if the area is stumpy and humidity is medium then production is medium but farmers received high level of yield.

Additionally, rule 61 described that the low area of cultivation with average humidity and production gives maximum rice yield. It is seldom happening and come to know with the rule number 64 that if the area is least and the other factors like wind speed and production is high and medium respectively then farmers can also get maximum rice yield. When the rule 66 was examined, it was observed that the area of cultivation may be low but if the wind speed is high and the production is average then we can obtain high level of yield. The farmers are upset due to the low area reserved for rice cultivation but rule number 73 defines that if the temperature and wind speed is high then yield will be high with the condition of medium production. Similarly, by applying rule number 74, the problem of area is same and the wind speed is high then it is possible to increase the temperature and yield except production.

Moreover, it is attractively observed with rule 76 that when the area is low but wind speed and temperature is at maximum level then yield is high but the production is medium. Another rule having number 77 is examined carefully and found that if area is low and temperature is high and production is medium then wind speed and yield both are high. Most of the farmers crop rice in low area due to unavailability of the proper land. They can get yield high when rule 79 is successfully executed. This rule holds good if wind speed is high, production is medium and temperature is high. We have also experimented rule number 84 that defines if the area is low and temperature as well as wind speed are high then yield is high but the production is medium. During the discussion with farmers and experts, it is found that low wind speed decreases the rice yield therefore with the consideration of rule 84 high level wind speed is beneficial for rice yield.

4 Summary and Conclusion

The aim of this study is to share yield predictions made on previous valuable rice data using analytical techniques and machine learning algorithms. The Apriori algorithm is implemented on

rice crop dataset of district Larkana. The required data is collected from the different associated departments. The association rule approach proved that the Rule 9, Rule 20, Rule 33, Rule 51 and others clearly results in "Yield is high". The predicted results of rice yield optimized using neural network provides high value of yield which can be followed by farmers, growers and government officials of Rice crop. STaleo that these personnels can take better decisions based on past trends and considering forecasting values of correlated attributes discussed in this study. The obtained results are bumper crop production that will strengthen country's economy by increasing rice production and also save the farmers from occurrence of monetary loss. The proposed neural network of optimization of rice yield of district Larkana will be analytically compared with the obtained results of advanced linear and statistical models.

References

1) Sellam V, Poovammal E. Prediction of Crop Yield using Regression Analysis. *Indian Journal of Science and Technology*. 2016;9(38):1–5. doi:10.17485/ijst/2016/v9i38/91714.
2) M U. Contribution of Agriculture Sector in the GDP Growth Rate of Pakistan. *Journal of Global Economics*. 2016;4(2). doi:10.4172/2375-4389.1000184.
3) Park JK, Das A, Park JH. Integrated Model for Predicting Rice Yield with Climate Change. *International Agrophysics*. 2018;32:203–215.
4) SUPRO IA, MAHAR JA, MAITLO A. Regression Analysis of Rice Data for Yield Prediction Using Python Programming Language. *SINDH UNIVERSITY RESEARCH JOURNAL -SCIENCE SERIES*. 2019;51(02):183–188. doi:10.26692/sujo/2019.6.32.
5) Surya P, Aroquiaraj IL, Kumar A. The Role of Big Data Analytics in Agriculture Sector: A Survey. *International Journal of Advanced Research in Biology Engineering Science and Technology*. 2016;2:830–838.
6) Rao PR, Gowda SP, Prathibha RJ. Paddy Yield Predictor Using Temperature, Rainfall, Soil pH, and Nitrogen. Emerging Research in Electronics. *Lecture Notes in Electrical Engineering 545*. 2019;p. 245–253.
7) Ramakrishna BV, Satyanarayana B. Agriculture Soil Test Report Data Mining for Cultivation Advisory. *International Journal of Computer Application*. 2016;6(2):11–16.
8) Supriya DM. Analysis of Soil Behavior and Prediction of Crop Yield using Data Mining Approach. *International Journal of Innovative Research in Computer and Communication Engineering*. 2017;5(5).
9) Zingade DS, Buchade O, Mehta N, Ghodekar S, Mehta C. Crop Prediction System using Machine Learning. *International Journal of Advance Engineering and Research Development*. 2017;4(5):1–6.
10) Chouhan S, Singh D, Singh A. A Survey and Analysis of Various Agricultural Crops Classification Techniques. *International Journal of Computer Applications*. 2016;136(11):25–30. Available from: https://dx.doi.org/10.5120/ijca2016908575. doi:10.5120/ijca2016908575.
11) Paswan RP, Begum SA. Regression and Neural Networks Models for Prediction of Crop Production. *International Journal of Scientific & Engineering Research*. 2013;4(9):98–108.
12) Pandey A, Mishra A. Application of artificial neural networks in yield prediction of potato crop. *Russian Agricultural Sciences*. 2017;43(3):266–272. doi:10.3103/s1068367417030028.
13) Niedbała G. Application of Artificial Neural Networks for Multi-Criteria Yield Prediction of Winter Rapeseed. *Sustainability*. 2019;11(2):533–533. doi:10.3390/su11020533.
14) Mathieu JA, Aires F. Using Neural Network Classifier Approach for Statistically Forecasting Extreme Corn Yield Losses in Eastern United States. *Earth and Space Science*. 2018;5(10):622–639. Available from: https://dx.doi.org/10.1029/2017ea000343. doi:10.1029/2017ea000343.
15) Khaki S, Wang L. Crop Yield Prediction Using Deep Neural Networks. *Frontiers in Plant Science*. 2019;10(621). doi:10.3389/fpls.2019.00621.
16) Gandhi N, Petkar O, Armstrong LJ. 2016.
17) Khaki S, Wang L, Archontoulis SV. A CNN-RNN Framework for Crop Yield Prediction. *Frontiers in Plant Science*. 2020;10. doi:10.3389/fpls.2019.01750.
18) Putri RE, Yahya A, Adam NM, Aziz A, S. Rice Yield Prediction with Respect to Crop Healthiness and Soil Fertility. *Food Research*. 2018;3(2).
19) Kuldeep S, Sunila, Sanjeev K. Crop Yield Prediction Techniques Using Remote Sensing Data. *International Journal of Engineering and Advanced Technology*. 2020;9(3):3683–3689.