Crop Prediction and Soil Analysis using Data Mining Techniques

Abstract - **Agriculture and its related industries holds a major contribution to our society as it relates to one of the basic needs of the human and in India it is the primary source of inhabitation for many people as it is a country of farmers. This research is aimed at farmers to improve their knowledge on sowing crops based on different classified soils. It will educate them on different factors like power of Hydrogen(pH), Electrical Conductivity (EC) and chemicals of the soil which will result in better yield of crops for the future. Research have been done using different statistical methods like Naïve Bayes, KNN, Decision tree etc. In this project, crop yield prediction is established by classification and prediction rule, where Principal Component Analysis, Support Vector Machine and Neural Networks techniques have been used.**

Keywords—soil analysis, crop prediction, data mining, support vector machine, principal component analysis, neural networks

I. INTRODUCTION:

Agriculture and agro-based industries play a major role as the primary source of inhabitation in India, which is often known as country of farmers. By gaining knowledge about soil, crops, factors influencing/affecting crops will help farmer in the long run. Also, as per the financial aspect, it will boost the economy of the country. Due to changes in weather conditions, farmers tend to change their occupation. So, its high time to analyze fast and beneficial crops depending on the soil [1]. And, this analysis on soil can be done with the help data mining. Soil, water, weather are the three important factors for soil cultivation. Soil analysis depends on various chemical properties such as Electrical Conductivity (EC), Power of Hydrogen (pH) with meso-elements Sulphur (S), macro-nutrients Zinc (Zn), Copper (Cu), Iron (Fe), Manganese (Mn) and micro-elements Nitrogen (N), Potassium (K) and Phosphorus (K) respectively.

Classification of soil plays a massive role in cultivation of crops. With the information about the previous work and with the help of data mining techniques, soil analysis problems will be reduced.

II. RESEARCH QUESTION:

Optimization of crop recommendation by implementing Data Mining techniques by Soil Analysis

Substantial amount of research has been done on yield prediction in agriculture domain. To determine type of crop which is suitable for a specific category of soil (properties and their range), pH and EC proved to be primary factor to predict target variable crop. The paper contains following sub-topics, (III) Literature Review of previous research on same domain with different machine learning algorithm, (IV) Methodology chosen for getting the output, (V) Implementation of machine learning algorithm for classification and prediction of attributes and (VI) Conclusion and future possible work which can be done for soil prediction.

III. LITERATURE REVIEW:

Author of [1] suggests farming has become low due to diverse climatic changes and low profit, so its high time to find some fast and beneficial methods for cultivation. This will motivate farmers to learn and grow more crops defending on that soil. Along with classification techniques and data mining algorithms such as Naive Bayes, JRIP and multivariate analysis was done to speed up the process. Although, this was only done on one crop Sunflower and this work can be extended to identify other suitable crops based that can be cultivated in that region based on their soil properties.

Author of [2] suggests genetic traits and climatic factors are the methods that can be determine potential crop yield. Understanding the pros and cons with the effects caused within soil and crop yields in an area can improve the crop yield effectivity. In this paper, CART (classification and regression trees) analysis was done. By this method, it predicts the different relationships and interactions between soil's chemical and physical properties to yield different classes. By improving soil properties of consistent areas helps to overcome the limiting factors of inconsistent yielding areas.

Author of [3] suggests, with the help of useful data mining techniques, agricultural sector will be benefit a lot. From a large dataset, the process of recapturing the information can be done with the help of data mining. Association Rule Mining, Clustering, Classification with the help of these 3 methods and various algorithms like Naive Bayes Classifier, J48, K-mean to solve agricultural problems. R programming was used to implement decision tree algorithm.

Author of [4] suggests prediction of crop is directly linked to various variables of soil quality. As, prediction of crop deals with large dataset, thus it is a perfect candidate for data mining techniques. To improve the performance of the prediction system, a robust and novel classifier needs to be created. With the help of nutrient contents, water tolerant of various seeds in case of flood and draught, sensors can be used to guide the farmers.

Author of [5] suggest classification techniques are important in soil analysis. It helps to classify sample of unknown dataset by applying info from classified sample. In K-Nearest Neighbor, it uses training set of data every time it is been performed. To use in assigning classification to unknown sample from the known sample the K parameter is used. On the other hand, Naïve Bayes classifier relates the presence or absence of a classifier depending on the accuracy of the model. It works way better in real scenario of complex situation.

Author of [6] suggested a model named crop yield prediction model (CRY), basically implies on adaptive cluster approach. It uses a modeling approach similar to bee hive to classify the

yield growth pattern. Using mining techniques likes of regression tree and clustering modelled in clementine and trained with large number of datasets of crop knowledge base. It works on mainly for data which is pre-processed making it easy to classify and cluster the crop based on crop yield. Performance of CRY is found to better if implemented with Clementine, future works should consider attributes such as environmental, social parameters and socio-economic attributes as constraints.

Author of [7] suggests, soil fertility is very important for cultivation. It depends on value of the crop obtained from sensor as well regional wise information about the crop. Predictive analysis is generally classified into two models: (a) Predictive and (b) Descriptive. NPK sensors were used which are portable and less time consuming, but it only detects NPK present in the soil thus giving it less accuracy on other nutrients. With the help of decision model, it was known that predictive model has few pros and cons. Main advantage been high accuracy ratio and it speeds up the process, while the main disadvantage was prediction based on atmosphere is not accurate every time as unpredictability, climate change and conditions may differ each time.

Author of [8] suggests climate change causes changes in the carbon balance of soil implying change on terrestrial carbon stock and the climate itself. To understand these changes and predict the future, model named as Earth system models (ESMs) is used. By using, Boosted Regression Trees (BRT) algorithm to identify the factors affecting the Soil Organic Carbon stock. This study also shows us how a data-mining algorithm can be used to assess different ESM model outputs.

Author of [9] suggests, there are various factors affecting farming process. With different parameters, a decision support system was created to assist farmers for selection of crop cultivation. Traditional method biggest challenge was selection of crops depending on the soil type. With the help of Data mining techniques like ANN user friendly android application was created but where there is less coverage of network, mobile application didn't work properly. Thus, new efficient techniques were needed.

Author of [10] suggests external factors such as rainfall, humidity, temperature, irrigated area and attributes such as soil salinity, soil pH plays a decisive role in crop cultivation. Data mining techniques were divided into 2 parts: (a) clustering and (b) classification. K-means clustering, neural net, linear regression, KNN was done using RapidMiner to predict RMSE values. Geospatial analysis was needed for more accuracy to yield different crops based on that area.

Author of [11] suggests, crop yield is directly influenced by the factors such as soil type, precipitation, seed quality etc. By using data mining techniques, crop yield prediction system was created. Classifiers like J48, Locally Weighted Learning (LWL), Logical Analysis of Data (LAD), Instance Based k-nearest neighbor (IBK) was compared using WEKA tool for accuracy. Although, this data was not based on real-time analysis. Real-time weather was next step of this project to check accuracy based on soil type.

Author of [12] suggests, soil quality plays a significant role in agriculture. In this paper, decision tree algorithm- C4.5 was used to predict model with the help of soil composition. C4.5 algorithm improves practical application, improves forecasting accuracy for precision fertilization. Real-time data was needed for the precision fertilization in agriculture for large agricultural soil data.

Author of [13] says, physical and chemical characteristics of soil can be solved by soil clustering and spatial representation. Data mining techniques like k-means clustering and fuzzy-k means clustering was done on the dataset of Montenegro and then was visualized on Google static map. Fuzzy k-means have a different degree of belonging, so the future work will define in a good manner of visualization.

Author of [14] suggests data mining and data visualization both are equally important to understand the trends of the effect of numerous factors influencing the crop yield. Decrease in precipitation increases in rice crop yield as well marginal changes in temperature also increases the rice crop yield. With the help of data mining techniques like J48, LAD Tree, IBk, LWL using WEKA. This paper showed various effects of climate scenarios and yearly trends in that selected region. Further, this method can be used to develop for other crops under that climatic scenario.

Author of [15] suggests information extracted from the database is used for decision making. Machine learning techniques like classification and prediction was used. Naïve Bayes classifier and Non-Naïve classifier was done. Non-Naïve classifier is an improved version of Naïve Bayes algorithm which removes the normal distribution of error probability of the distribution. Soybean dataset was used for analysis. Future work on this project is how a user can interact on queries like soil type, tillers, seeding etc.

Author of [16] suggests use of decision tree methodology to a classified set of data gets the key interpretations out of it. It can easily differentiate between different classified labels of data and predict it accurately. Using multiple attributes by object not going to be effective for classification. More corelated attributes are selected by algorithm and it makes the prediction tree around it as they are more specific avoiding lack of generalization. Using multi-attributes by object did not have much contribution to the classification improvement. This project depends on mean pixel values which has a strong correlation with attributes of the selected region, thus lacking in other landscape situations.

Author of [17] suggest information extracted from database is used for decision making. To get insights, trends and predictions for a dataset, classification and prediction algorithm of machine learning is being used. Classification used for labeled or distinct values whereas prediction is used for continues values. Naïve Bayes classifier is a model which is based on probability assuming one attribute is fully independent of other’s value which is called as class conditional independence. The independent model uses underlying probability and uses generative algorithm applying Bayesian rule and the independent assumption. It takes less time to get the prediction for large data set making it obvious choice for most predictions. Speaking about Non- Naïve classifier, which is an improved version of original Naïve Bayes algorithm which removes the normal distribution of error probability of the distribution. This makes it faster and more efficient as co pared to normal Bayesian algorithm.

IV. METHODOLOGY:

Data Mining is an essential process that is included in KDD methodology. KDD is known as Knowledge Discovery in Databases. The vital steps in KDD are:

1) Data Cleaning: Cleaning of unnecessary, null values was done.
2) Data Integration: Multiple data sources may be combined.
3) Data Selection: Relevant data for the analysis was retrieved from the database.
4) Data Transformation: Data was transformed and buildup into appropriate forms for data mining.
5) Data Mining: Which is most important step where proper data mining algorithms to be implemented to extract data patterns.
6) Data Evaluation: After successful evaluation of different patterns by data mining, interpretation of results is done.
7) Knowledge Presentation: Representation techniques with the help of visualization is done for reporting and decision making.

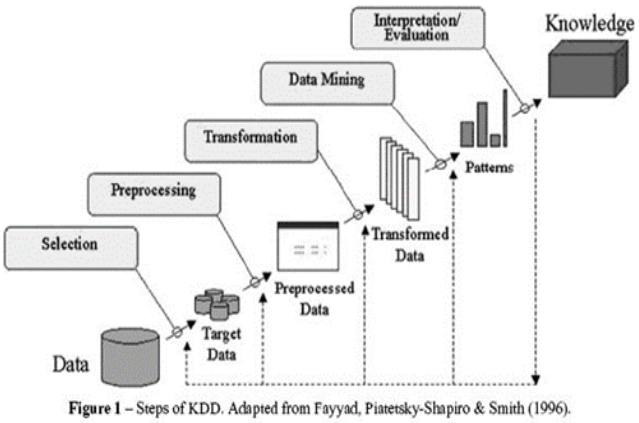

Fig. 1 - Steps of KDD

a) Target Dataset Selected:

The data have been taken from National Bureau of Soil Survey and Land Use Planning website from the publication of data of Tumkur district. The data was in book format and extracted using excel. It has many important attributes which contributes to the class of soil. After that by using classification done in for different groups depending on their values of the soil properties [18].

b) Attributes:

Place ID (Place ID of regions in Tumkur District, India), pH (power of hydrogen), EC (Electrical Conductivity), OC (Organic Carbon), N (Nitrogen), P (Phosphorus), K (Potassium), Zn (Zinc), Cu (Copper), Fe (Iron), Mn (Manganese), S (Sulphur), Crop which is a target variable.

c) Exploratory Data Analysis:

The exploratory data analysis is most important techniques for Data Mining. The null or bad data were cleaned through Microsoft Excel and R programming.

d) Removing inappropriate attributes:

Once the analysis and cleaning are done, the next step is to select the needed attributes leaving the in affecting ones. For instance, the place id which does not contribute to research has been left out from the implementations.

V. IMPLEMENTATION:

PRINCIPAL COMPONENT ANALYSIS:

Principal component analysis is used for selecting important factors and variables from a large set of attributes of a larger dataset, which leads to concentrate on the factors which are important rather than evaluating all the available variables.
It separates the variables with respect to their low and high effect on the dependent variable providing more information about the relations.
It is easy to understand and interpret the data meaningfully and to visualize it. This method is useful when the data is having more than 2-dimensional data variable.
To apply the principal component analysis, we need to have all numerical values in our dataset. To do so, we have classified the crop column to different numeric group of 1 to 4 depending upon their group of crops. It makes the dataset numeric and standardize to do the calculations.
A principal component is a normalized linear combination of the original predictors in a data set.
principal component is linear combination of normalized dataset of predictors.

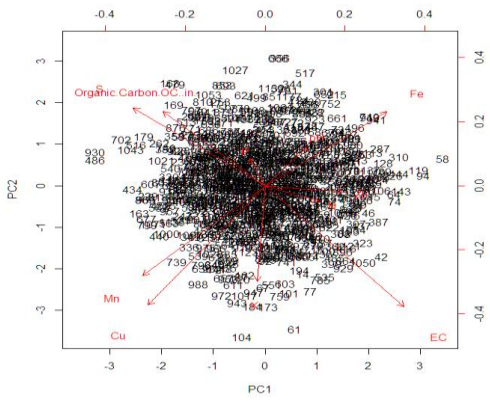

Fig. 2 – symmetry of PC1 and PC2

In the above image, *PC1* and *PC2* are the symmetry of respective principal components which are been obtained from our dataset.
The mathematical formulae for principal component:
$Z^1 = \Phi^{11}X^1 + \Phi^{21}X^2 + \Phi^{31}X^3 + ... + \Phi^{p1}X^p$
Where,
X^1, X^2..., Xp are set of normalized predictors having mean of zero and standard deviation of one.
Similarly, we can compute the second principal component as
$Z^2 = \Phi^{12}X^1 + \Phi^{22}X^2 + \Phi^{32}X^3 + ... + \Phi^{p2}X^p$
If the two components are uncorrelated, their directions should be orthogonal. The direction of components should

be orthogonal if there is no correlation exits between them.
From the measurement of variance, we got proportion of variance which explains that the first principle of component explains 10.4% of variance, second one explains 10.2%, third one explains 9.9% of variance and so on. To decide which of the factors to consider we have gone ahead to plot scree plot and check what it suggests.

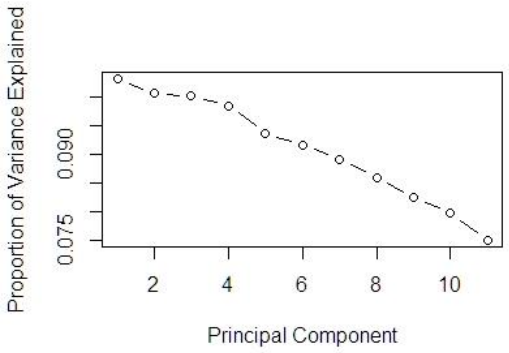

Fig.3 – Scree Plot of Principal Component vs. Proportion of Variance

From this scree plot, we can see that all our components are needed as everything is in between 70% to 90% of variance in data set. With the root node of 914 and split correlation of 1966.927 and a deviance of 3.020011, PCA variance was in the good range.

NEURAL NETWORK:

Neural networks comprise of interconnected entities which are highly correlated. These entities are called nodes or units [14]. In neural networks there is a concept of perceptron which is now called as Artificial neuron. It is anything which accepts two or more inputs processes it and gives the output. Neuron accepts inputs which are weighted accordingly to the desired output. Neural networks algorithm is inspired from the biological nervous system [14]. Neural network algorithms consist of a learning procedure through which it is used to classify and predict the output. For present research the multilayer perceptron was used as neural network for the prediction of crop and soil analysis yield for Tumkur district in Karnataka state, India.

Neural networks use a non-linear transformation for the inputs. It is also called activation function. What activation function does is it takes the sum of all the weighted input and return the output of that neuron.
Single layer perceptron for simplicity is used to explain the concept of neural networks. But for practical purposes we use Multilayer Perceptron. Multi-layer perceptron consists of many hidden layers. These are situated in between Input and Output layers.

Neural network is adaptive in nature which means that it learns itself from the source and train itself to optimize the weights for a better prediction with unknown situations.

$$a = f\left(\sum_{i=0}^{N} WiXi\right)$$

Activation Function takes the sum of weighted input (w1*x1 + w2*x2 + w3*x3 + 1*b) as an argument and return the output of the neuron. In above formula, i as x0 and b as w0.

The activation function is mostly used to make a non-linear transformation which allows us to fit nonlinear hypotheses or to estimate the complex functions

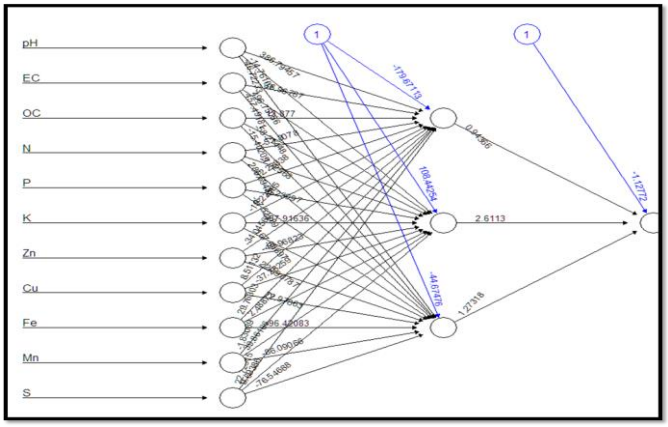

Fig. 4 – Neural network

In this project we have used neural network to predict the crop by analyzing the minerals of the soil. The following picture shows us the input layer, output layer and the hidden layer which consists of three perceptron's which has been explained earlier. The black lines depict the weights of each element and which is nothing but their contribution to the output and the blue lines shows the bias. Bias helps us to analyze how flexible a perceptron can be. It can also be considered as an intercept used in the line equation y = mx + b where b is the intercept. Bias is used to move up or move down the line equation to fit the model and predict the accurate data. Split covariance (or correlation) matrix into scale part (eigenvalues) and direction part (eigenvectors).

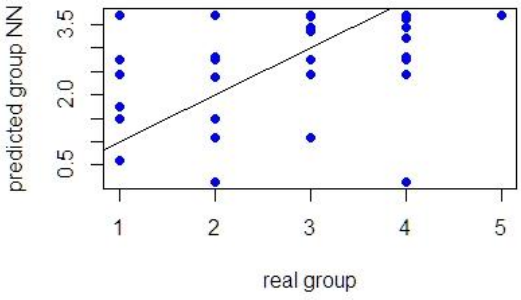

Fig.5 – Real Group vs Predicted Group NN

Deviance as a function of iteration. For logistic PCA, the deviance should decrease at each iteration. The Root Mean Square Error (RMSE) is used to calculate the difference which exist between the measured value of predictive model and the actual value of the dataset. The RMSE value is 0.94 for our model. As our dataset is small and the value of RMSE is 0.94 which is between 0 and 1 therefore it is good value for our model.

SUPPORT VECTOR MACHINE:

Support Vector Machine (SVM) is a Machine Learning Algorithm tool used for classification and regression which is used to solve the real-time issues to give a good data performance. The Parameters in SVM are the best fit to solve the experimental issues in an architecture of the learning machine. It is a classification technique.
Support Vector Machine is like a sharp knife - it works on smaller datasets, it can be much more strong and powerful in building models. The basic idea of Support Vector Machine is to identify the space between the data points from the given data. Splitting of data is the aim of this algorithm in which it finds the best possible way to split the data. The data points are called as support vector and the splitting line is called as hyperplane. The distance between the support vector and the hyperplane are as far as possible. The job of SVM is trying to maximize the margin.

a) Overall Model:

This result shows the overall statistics of the Crop:

```
Overall Statistics

               Accuracy : 0.4743
                 95% CI : (0.4429, 0.5059)
    No Information Rate : 0.4743
    P-Value [Acc > NIR] : 0.5125

                  Kappa : 0
 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 1 Class: 2 Class: 3 Class: 3,4 Class: 4
Sensitivity            0.0000   0.0000   0.0000    0.00000   1.0000
Specificity            1.0000   1.0000   1.0000    1.00000   0.0000
Pos Pred Value            NaN      NaN      NaN        NaN   0.4743
Neg Pred Value         0.8892   0.8922   0.7382    0.95468      NaN
Prevalence             0.1108   0.1078   0.2618    0.04532   0.4743
Detection Rate         0.0000   0.0000   0.0000    0.00000   0.4743
Detection Prevalence   0.0000   0.0000   0.0000    0.00000   1.0000
Balanced Accuracy      0.5000   0.5000   0.5000    0.50000   0.5000
```

Fig.6 – Overall Crop Statistics

The accuracy of the Crop by predicting it with the pH and EC value is 47.4% and the calculated probability P-Value [Accuracy > NIR] is 51.25 %. The Kappa co-efficient (k) is a statistical measure for inter-rater agreement for categorical items here is 0, which indicates that the degree of agreement is as expected by chance. The Sensitivity P(+|D) and Specificity P(-|D) shows the True positive and True Negative and from the observation we can understand that the Crop 4 has the highest positive range of higher growth and the remaining crops 1, 2, 3 and 3,4 (correlation of crops 3 and 4) has the non-diseased crops getting negative results as they have possibility of less crop growth in the soil according to the pH and EC value. Positive Prediction value of the crops 1,2,3 and (3,4) has no diseased value, as their sensitivity results in 0 and the crop 4 has 47% chance of disease. At the same time Negative prediction value of the crops 1,2,3 and 3,4 are more than 70%.

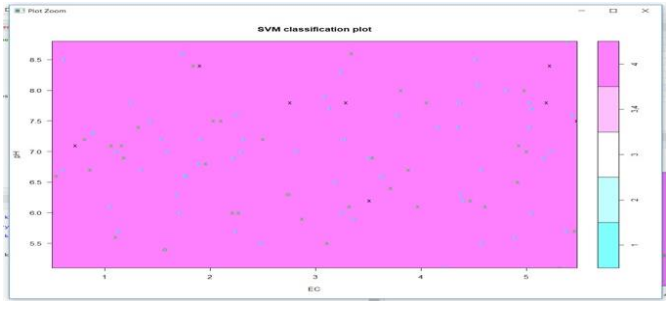

Fig. 7 – SVM Classification Plot

The SVM classification plot shows that the data points for the crops are graphed between pH value and EC value. The data points are the Crop values which are randomly selected from the data set. The data points in this model are overlapped with each other since there is a correlation between them and due to the highest positive rate, the crop 4 dominates the plot so that it has less chance to draw the hyper plane. The data points of the crops were indicated by different colors of x and 0. Thus this plot of SVM explains that the Crops in the range 4 has the highest possibility to grow according to the pH and EC value in the soil.

b) Chemicals:

In this project, the SVM algorithm plays a major role to develop the predictive modelling which gives the accuracy of the chemicals of each range (Low, Medium and High) found in the soil. This accuracy gives us the detailed prediction about the best crop to grow on a suitable soil.

The dataset deals with 1143 records of 12 columns which is loaded to create the Training and Test samples. From this record the column 1 which has the Place.ID is not required for detailing the prediction, so it is eliminated from the testing and training datasets. The columns 4,5,10 are N(Nitrogen), P(Phosphorous), and Mn(Manganese) respectively has only Medium range with them so it has no meaning in it to find the accuracy as they have proven that they have 100% of Medium accuracy in them. The remaining chemicals Zn, Cu, S, Fe and K needs to be calculated to find the accuracy. 70 % of data were randomly sampled for training and the remaining 30% for testing.

Measure/ Components	Copper	Zinc	Sulphur	Iron	Potassium
Accuracy	0.3947	0.52	0.4708	0.6696	0.5322
Kappa value	-0.0.241	0.064	-0.0.489	0.257	0.0724
Mcnemar's Value	<2e-16	2.53e-10	1.96e-07	0.1323	1.18e-07
Sensitivity	0.363	0.4938	0.4706	0.5429	0.5149
Specificity	0.58	0.5859	0.4712	0.7257	0.5701
Prevalence	0.8538	0.7105	0.6959	0.307	0.6871

Fig.8 – Overall Chemical Statistics

The caret (Classification and Regression Training) package is installed in this program which can tune SVM algorithm which is the attempt to create the predictive models.

The tools available in the package for:

- Data Splitting.
- Feature selection.
- Preprocessing.
- Model resampling.
- Variable estimation.

The tuning parameter C is used in the Grid section during down sampling the train control function. The three kernels Linear, Radial and Polynomial in SVM were implemented to compare the resamples of data to find the ROC, Sensitivity and Specificity of various values of the chemicals find in the grid. Multiple models were tuned to find the best fit value for ROC.

VI. CONCLUSION:

In this project, different models of Data Mining have been explored to study yield production depending on soil type. Soil is been classified from low to high category depending upon their attributes using mining techniques. Principal Component Analysis have been used to check important factors affecting the target variable. Neural Network has found to significantly perform with a good RMSE value which is the moderate range, while Support Vector Machine is a classification and regression method, predict the crop range according to soil properties based in that region.

The future work is to create more efficient data mining models and checking other chemicals and factors like rainfall, temperature, humidity etc. of the soil for better crop production of other regions.

REFERENCES

[1] Wang, Y.P. and Shen, Y., 2015. Identifying and characterizing yield limiting soil factors with the aid of remote sensing and data mining techniques. Precision agriculture, 16(1), pp.99-118.

[2] Karthik, D. and Vijayarekha, K., 2015. Simple and Quick Classification of Soil for Sunflower Cultivation using Data Mining Algorithm. International Journal of Chem Tech Research, 7(6), pp.2601-2605.

[3] Hemageetha, N., 2016, March. A survey on application of data mining techniques to analyze the soil for agricultural purpose. In Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on (pp. 3112-3117). IEEE.

[4] Gandge, Y., 2017, December. A study on various data mining techniques for crop yield prediction. In Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), 2017 International Conference on (pp. 420-423). IEEE.

[5] Paul, M., Vishwakarma, S.K. and Verma, A., 2015, December. Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach. In Computational Intelligence and Communication Networks (CICN), 2015 International Conference on (pp. 766-771). IEEE

[6] Ananthara, M.G., Arunkumar, T. and Hemavathy, R., 2013, February. CRY—an improved crop yield prediction model using bee hive clustering approach for agricultural data sets. In Pattern Recognition, Informatics and Mobile Engineering (PRIME), 2013 International Conference on (pp. 473-478). IEEE.

[7] Vijayabaskar, P.S., Sreemathi, R. and Keertanaa, E., 2017, March. Crop prediction using predictive analytics. In Computation of Power, Energy Information and Commuincation (ICCPEIC), 2017 International Conference on (pp. 370-373). IEEE

[8] Hashimoto, S., Nanko, K., Tupek, B. and Lehtonen, A., 2017. Data-mining analysis of the global distribution of soil carbon in observational databases and Earth system models.

[9] Shirsath, R., Khadke, N., More, D., Patil, P. and Patil, H., 2017, June. Agriculture decision support system using data mining. In Intelligent Computing and Control (I2C2), 2017 International Conference on (pp. 1-5). IEEE.

[10] Ahamed, A.M.S., Mahmood, N.T., Hossain, N., Kabir, M.T., Das, K., Rahman, F. and Rahman, R.M., 2015, June. Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh. In 2015 16th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD) (pp. 1-6). IEEE.

[11] Mishra, S., Paygude, P., Chaudhary, S. and Idate, S., 2018, January. Use of data mining in crop yield prediction. In 2018 2nd International Conference on Inventive Systems and Control (ICISC). IEEE.

[12] Dongming, L., Yan, L., Chao, Y., Chaoran, L., Huan, L. and Lijuan, Z., 2016, October. The application of decision tree C4. 5 algorithm to soil quality grade forecasting model. In Computer Communication and the Internet (ICCCI), 2016 IEEE International Conference on (pp. 552-555). IEEE.

[13] Hot, E. and Popović-Bugarin, V., 2015, November. Soil data clustering by using K-means and fuzzy K-means algorithm. In Telecommunications Forum Telfor (TELFOR), 2015 23rd (pp. 890-893). IEEE.

[14] Gandhi, N. and Armstrong, L.J., 2016, October. Rice Crop Yield forecasting of Tropical Wet and Dry climatic zone of India using data mining techniques. In Advances in Computer Applications (ICACA), IEEE International Conference on (pp. 357-363). IEEE.

[15] Santhosh, D., Pandiyaraju, V. and Kannan, A., 2014, December. Non-Naive Bayesian classifier for Farmer Advisory System. In Advanced Computing (ICoAC), 2014 Sixth International Conference on (pp. 248-254). IEEE.

[16] Goltz, E., Arcoverde, G.F.B., de Aguiar, D.A., Rudorff, B.F.T. and Maeda, E.E., 2009, July. Data mining by decision tree for object-oriented classification of the sugar cane cut kinds. In Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009 (Vol. 5, pp. V-405). IEEE.

[17] Santhosh, D., Pandiyaraju, V. and Kannan, A., 2014, December. Non-Naive Bayesian classifier for Farmer Advisory System. In Advanced Computing (ICoAC), 2014 Sixth International Conference on (pp. 248-254). IEEE.

[18] Arooj, A., Riaz, M. and Akram, M.N., 2018, February. Evaluation of predictive data mining algorithms in soil data classification for optimized crop recommendation. In Advancements in Computational Sciences (ICACS), 2018 International Conference on (pp. 1-6). IEEE