© 2017 IJSRST | Volume 3 | Issue 3 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

Expert System for Crop selection

Somnath Devadhe, Abhijit Kausadikar, Pratik Daphal, Akshay Joshi
Department of Computer Engineering, KKWIEER, Nashik, Maharashtra, India

ABSTRACT

Despite technological advancements, farming sector in India remains un-organized. Agriculture of India needs essential changes with respect to its current social, geographic and economic trends by the combination of practical knowledge gathered over generations and the scientific basis. There is a need to adapt artificial intelligence in agricultural sector due to its promising results in fields ranging from medical science to automated machines. With the help of artificial intelligence, computations on historical data can be performed to predict crop productions. To predict the productivity of crops, this paper demonstrates the use of various machine-learning techniques such as linear regression, decision tree and random forest. Data visualization techniques are used to present region wise patterns of crop productivity using HTML5, AngularJS etc.

Keywords: Artificial intelligence, Visualization, Machine learning

I. INTRODUCTION

Nowadays, Artificial intelligence has emerged as vivid technology to reduce human efforts and errors by perceiving environment and taking actions accordingly. Machine learning is a core sub-area of artificial intelligence as it enables computers into training and testing methodology.

In agriculture-based countries, economic growth and food security significantly depends upon agriculture planning. Crop selection is an important issue for agriculture planning. It depends on various parameters such as crop yield, market value and government policies. Many researchers studied prediction of yield rate of crop, prediction of weather, soil classification and crop classification for agriculture planning using statistics methods or machine learning techniques. Selection of crop is a puzzle when there is more than option of crop to be cultivated in limited resources.

Machine learning is the ability to learn from examples. Machine learning algorithms consist of two types: unsupervised learning and supervised learning. Unsupervised learning is a cluster analysis to find hidden pattern by exploratory data analysis. Supervised learning tries to find a relation between a subset of the attributes, called inputs or attribute variables, and the dependent attribute or outputs. Classification and regression are the two approaches for supervised learning. The predicted output for classification and regression is categorical and numerical respectively. In this case, the concerned output is numerical. Hence, various regression techniques are considered.

DISTRICTWISE AREA ,PRODUCTION & PRODUCTIVITY OF PRINCIPAL CROPS DURI
Area in "00" ha, Production in "00" Tonnes, Productivity in Kg /ha.

District	Kh Rice			Summer Rice			Total Rice		
	A	P	Py	A	P	Py	Area	Prodn	Prodty
Mumbai									
Thane	1393	1702	1222	24	51	2136	1417	1753	1237
Raigad	1264	2741	2168	83	176	2119	1347	2917	2166
Ratnagiri	784	1659	2116	4	12	2914	788	1671	2121
Sindhudurg	754	1873	2484	44	102	2323	798	1975	2475
Konkan Dn.	**4195**	**7975**	**1901**	**155**	**341**	**2200**	**4350**	**8316**	**1912**
Nasik	456	398	873	0	0	0	456	398	873
Dhule	49	18	367	0	0	0	49	18	367
Nandurbar	210	49	233	0	0	0	210	49	233
Jalgaon	17	21	1176	0	0	0	17	21	1176
Nasik Dn.	**732**	**486**	**664**	**0**	**0**	**0**	**732**	**486**	**664**
Ahmednagar	55	38	691	0	0	0	55	38	691
Pune	640	752	1175	0	0	0	640	752	1175
Solapur	30	34	1138	0	0	0	30	34	1138
Pune Dn.	**725**	**824**	**1137**	**0**	**0**	**0**	**725**	**824**	**1137**
Satara	439	513	1169	0	0	0	439	513	1169

District	May			June			July
	Normal Ra	Actual Ra	Rainy Day	Normal Ra	Actual Ra	Rainy Day	Normal Ra
Thane	16.5	0.9	0	455.5	569	18	1003.6
Raigad	25.5	26.8	3	650.6	546.2	21	1181.2
Ratnagiri	36.4	88.5	11	817.9	544.2	22	1286.4
Sindhudu	64.1	83.3	9	913.8	524.8	20	1254.5
Palghar	15.6	0	0	426.6	747.8	19	1047.8
Nasik	19.1	1	0	154.4	229.5	15	378.1
Dhule	10.7	0	0	116.7	188.1	10	168.7
Nandurba	8.5	0	0	123.8	351.8	10	305.9
Jalgaon	10.9	0.7	0	130	168.1	10	206.8
Ahmedna	19.9	3.4	1	101.4	88.9	8	102.5
Pune	29.8	3	0	139.9	140	10	286.6
Solapur	25.6	13.4	2	102	56.3	5	101.1
Satara	39.2	27	4	149.3	152.8	15	339.9
Sangli	43.2	13.3	2	85.2	30.3	3	122.3
Kolhapur	70	26.3	4	337.8	297.1	18	757.4
Aurangab	16.3	0	0	131.4	98.9	9	168.1
Jalna	17.2	0	0	138.9	151.2	9	171.8

Figure 1: samples of data collected from government portals.

In figure 1, the independent values are in the form of rainfall and dependent values in crop productivity as the productivity of crops is dependent on the rainfall.

II. RELATED WORK

In [1], an agricultural expert system is proposed to suggest a structural view to increase production, produce less pesticide-infected food and conserve nature. This AES mainly consists of Crop Selection Expert System (CSES) and Disease Management Expert System.

Indian agriculture is highly dependent on summer rainfall. The correlation between summer rainfall and agriculture product production is studied in [4]. This paper presents an analysis of crop-climate relationship using past crops.

[2]This paper proposed a method named Crop Selection Method (CSM) to solve crop selection problem, and maximize net yield rate of crop over season and subsequently achieves maximum economic growth of the country.

To resolve uncertainty of agriculture production, dynamic dependence and knowledge ambiguity, [3] uses computer modelling and visualization techniques combined with artificial intelligence, intelligent information systems, database and decision support systems.

III. PROPOSED WORK

This paper proposes use of various machine learning techniques to increase accuracy to predict the yield rate of crops that can be considered in different phases of crop planning such as plantation, cultivation etc.

While dealing with the predictions of crop productions, various statistical and informative data about crops is considered. Information such as district wise monthly rainfall occurred over the years and district wise production of crops during the consecutive years is easily available on government portals. This data needs to be pre-processed and cleansed so that it can be used for training and testing to the expert system.

Indian agriculture is mainly dependent on rainfall. So, the relation between the rainfall and crop productions is established to predict the future productions of crops. The rainfall from month of June to November is considered. The values of crop productivity and rainfall are in kg/ha and mm respectively. Visualization of statistical data available and predictions w.r.t the statistical data is made to showcase the crop scenarios.
Machine learning consists of training and testing of data to the system. So a ratio of 80:20 is used to train and test the data. District wise monthly rainfall and productivity of crops during 2000 to 2014 in Maharashtra are considered. Details of the various machine-learning techniques used are as follows:

Linear Regression

Linear regression tries to define the relationship between two variables by fitting a linear equation to the observed data. One variable is considered an explanatory variable, and the other is considered a dependent on the explanatory variable. A modeller should first check if there is a relationship between the variables of interest, before attempting to fit a linear model to observed data. Though this cannot be termed as one variable causes the other, there is some notable association between the two variables. $\mathbf{Y = p + qX}$ is the form of equation linear regression line has, where **X** is the explanatory variable and **Y** is the dependent variable. The slope of the line is **q**, and **p** is the intercept (the value of **y** when $\mathbf{x} = 0$) [5].

Decision tree

Decision tree learning splits entire sample space recursively into smaller sub-sample space which is enough to be formulated by a simple model [8]. The ultimate goal of sub-sample using decision tree method is to mitigate mixing of different outputs values and assign single output value for sub-samples space. A decision node has two or more branches each representing values for the attribute tested. Leaf node represents a decision on the numerical target [6]. The topmost decision node in a tree that corresponds to the best predictor called root node. Decision trees can handle categorical as well as numerical data. As the target variable does not have classes in regression tree, a regression model is fitted to the target variable using each of the independent variables.

Random Forest

Random Forest is a bootstrap aggregating technique, which is based on tree ensemble machine learning method [9]. It creates multiple trees of randomly sub-sampled features. The output of forest is evaluated by considering average value of the prediction of individual trees. Random Forest can be used in high dimension input predictor as it has random sub-sampled features. Random forests train a set of decision trees separately, so the training can be done in parallel. The algorithm adds randomness into the training process so that each decision tree is a bit different. Combining the predictions from each tree reduces the variance of the predictions, improving the performance on test data [7]. In random forest regression, averaging takes place where each tree predicts a real value and the label is predicted to be the average of tree predictions.

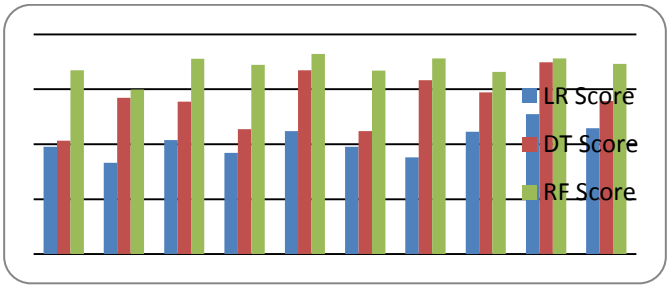

Figure 2: Comparison chart of cross validation score for predictions by linear regression, decision tree and random forest.

The comparison for prediction accuracy between linear regression, decision tree and random forest is shown in figure 2. It is observed that the decision tree method gives better predictions than linear regression. Random forest method performs better than the decision tree.

IV.CONCLUSION

This paper presents the use of various machine-learning techniques to predict yield rate of crops which will help select the sequence of crops to be planted over season. This may improve the accuracy of predicted yield rate of crops. The proposed system resolves the issue of selection of crop(s) based on yield rate prediction by using artificial intelligence to compare historical data for intelligent decision-making. This system successfully combines visualization techniques and computer modelling with artificial intelligence, intelligent information systems. It is found out that the random forest method gave better prediction results. Other parameters such as water density, soil type, crop type can be considered of crops not cultivated in rainy season for prediction of crop productivity in future scope.

V. REFERENCES

[1]. Balmukund Maurya; Mohd Rizwan Beg; Sudeep Mukherjee, Expert system design and architecture for farming sector, Information & Communication Technologies(ICT), 2013 IEEE Conference on, Year: 2013

[2]. Rakesh Kumar; M. P. Singh; Prabhat Kumar; J. P. Singh, Crop Selection Method to maximize crop yield rate using machine learning technique, Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), 2015 International Conference on Year: 2015

[3]. Liying Cao, Zhang Xiaoxian, Yu helong1, Chen Gui-fen,The rule model and visualization technology application in intelligent decision-making system, Second International Conference on Instrumentation & Measurement, Computer, Communication and Control, 2011

[4]. K. Krishna Kumar, K. Rupa Kumar, R. G. Ashrit, N. R. Deshpande and J. W. Hansen, ”Climate Impacts on Indian Agriculture”. International Journal of climatology, 24: 13751393, 2004.

[5]. Linear Regression, http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm

[6]. Decision tree – regression, http://www.saedsayad.com/decision_tree_reg.htm

[7]. Ensembles – RDD based API, https://spark.apache.org/docs/latest/mllib-ensembles.html#random-forests

[8]. Quinlan JR, ”Learning with continuous classes”. Proc. AI92, 5th Aust. Joint Conf. on Artificial Intelligence (Adams & Sterling, eds.), World Scientific, Singapore, pp: 343-348, 1992.

[9]. L. Breiman, ”Random forests”. Machine Learning, vol. 45, no. 1, pp. 532, 2001.