

Agricultural Analysis Using Data Mining And Machine Learning

G.Elizabeth Rani¹, N.Pooja²

¹Assistant professor, Dept of Computer Science and Engineering

²Dept of Computer Science and Engineering

^{1,2} Kalasalingam Academy of Research and Education, Anand Nagar,
KrishnanKoil, India

Abstract- Agriculture is the most importance in India. The latterday technologies can change the situations of farmers and decision makers in agricultural field. To analyze the dataset of agriculture, Python is used to predict the crop production. We used Jupyter Notebook which is a data mining tool to predict the crop production. In FAOSTAT dataset contains the parameter are precipitation, temperature, reference crop, area, evapotranspiration, production and yield for the season from January to December for the years 2000 to 2018. The data mining techniques are like K-means clustering, KNN, Bayesian network algorithm and SVM are used. Using these algorithms, where high accuracy can be achieved. Precision agricultural is a key component. Precision Agriculture Classification on FAOSTAT Dataset is perform using Artificial Neural Networks.

Keywords- Agriculture analysis, Data mining, K-means clustering Algorithm, SVM, KNN.

I. INTRODUCTION

Agriculture is one of the major source in Indian Economy. Day by day, the population increases, So that the demand of the food also increases. To get cleanse of these situations of farmers, researchers and agricultural scientists are trying to get better crop yield. Data mining is the process of analyzing or discovering the hidden patterns in large data sets, according to various perspectives for classification and its converted into relevant information. Data's are arranged in particular areas like data repository. This systematic analysis using data mining techniques helps farmers to take decisions. This details or information help them to reduction of costs and increasing the production rate. There are four steps followed in data mining process, they are extracting, transforming and loading he data in a repository, Then managing the data in multidimensional databases. Using application software, data mining provides data access to analysts. The analyzed data is represented using graphs.

II. LITERATURE REVIEW

Agriculture is one of the major source in Indian Economy. Day by day, the population increases, So that the demand of the food also increases. To get cleanse of these situations of farmers, researcher and agricultural scientists are trying to get better crop yield, analyzing process of hidden patterns according to various perspectives for classification and converted into relevant information is called as data mining in which data is arranged in particular areas like data repository. The efficient analysis using data mining techniques help farmers to take decisions. These information help them to reduction of costs and increasing the production rate. This method used the Food and Agriculture Organization of the United Nations (FAOSTAT) dataset which can be used as the open source. After collecting data, the dataset will be pre-processed to reduce the unwanted data in order to make data can be readable in the next step. Dataset will be separated into three parts: Train, Validation, and Testing data. ANN algorithm used to predict the crop for year of production. In the first phase, data frame are read and processed. Then it set the list of crop per season, temperature and by rainfall. ANN matrices to predict state for crop dataset are applied in the second phase. The regression format dataset calculated by ANN and gives result.

III. EXISTING WORK

Data mining techniques like k-means, k nearest neighbor, SVM, and Bayesian network are used. In k-means, a qualitative method of splitting a group of data. A quantitative approach is used to measure unique characteristics of the products.

IV. EXPERIMENTAL RESULT AND DISCUSSION

Data collection and preprocessing

For our process, data's are collected from agriculture department. In FAOSTAT, the dataset contains the parameter are Precipitation, reference crop, temperature, area, production, evapotranspiration and yield for the season from January to December for the years 2000 to 2018. The unwanted data's are removed from the dataset, then data's are uploading into the database in preprocessing.

YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
1995	42.3	37.7	25.0	0.0	328.0	337.0	304.9	481.1	550.0	554.0	555.2	534.0	5372.1
1996	57.0	6.1	12.0	0.0	444.1	337.1	228.9	750.7	688.2	717.2	328.1	169.9	3225.7
1997	12.7	14.4	0.0	0.0	238.1	476.0	228.4	326.7	329.0	234.4	225.0	267.0	2657.4
1998	8.4	14.4	0.0	0.0	322.0	368.0	400.1	382.0	500.0	422.0	230.0	100.0	3679.5
1999	1.8	0.0	3.0	26.0	278.0	328.7	380.7	320.0	287.0	207.0	28.4	44.4	2848.7
2000	30.0	0.0	0.0	0.0	100.0	720.0	247.7	320.0	104.0	217.0	120.0	79.2	2204.4
2001	21.0	21.0	816.0	305.0	440.0	377.0	200.0	424.0	424.0	324.0	324.0	324.0	3240.0
2002	10.0	10.0	0.0	0.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	1000.0
2003	20.0	20.0	20.0	20.0	20.0	20.0	20.0	20.0	20.0	20.0	20.0	20.0	200.0
2004	30.0	30.0	30.0	30.0	30.0	30.0	30.0	30.0	30.0	30.0	30.0	30.0	300.0
2005	40.0	40.0	40.0	40.0	40.0	40.0	40.0	40.0	40.0	40.0	40.0	40.0	400.0
2006	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	500.0
2007	60.0	60.0	60.0	60.0	60.0	60.0	60.0	60.0	60.0	60.0	60.0	60.0	600.0
2008	70.0	70.0	70.0	70.0	70.0	70.0	70.0	70.0	70.0	70.0	70.0	70.0	700.0
2009	80.0	80.0	80.0	80.0	80.0	80.0	80.0	80.0	80.0	80.0	80.0	80.0	800.0
2010	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	900.0
2011	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	1000.0
2012	110.0	110.0	110.0	110.0	110.0	110.0	110.0	110.0	110.0	110.0	110.0	110.0	1100.0
2013	120.0	120.0	120.0	120.0	120.0	120.0	120.0	120.0	120.0	120.0	120.0	120.0	1200.0
2014	130.0	130.0	130.0	130.0	130.0	130.0	130.0	130.0	130.0	130.0	130.0	130.0	1300.0
2015	140.0	140.0	140.0	140.0	140.0	140.0	140.0	140.0	140.0	140.0	140.0	140.0	1400.0
2016	150.0	150.0	150.0	150.0	150.0	150.0	150.0	150.0	150.0	150.0	150.0	150.0	1500.0
2017	160.0	160.0	160.0	160.0	160.0	160.0	160.0	160.0	160.0	160.0	160.0	160.0	1600.0
2018	170.0	170.0	170.0	170.0	170.0	170.0	170.0	170.0	170.0	170.0	170.0	170.0	1700.0
2019	180.0	180.0	180.0	180.0	180.0	180.0	180.0	180.0	180.0	180.0	180.0	180.0	1800.0
2020	190.0	190.0	190.0	190.0	190.0	190.0	190.0	190.0	190.0	190.0	190.0	190.0	1900.0

Training Process

In Training process, to predict the result, train the dataset. Artificial Neural Network method is applied to train the data prediction crop for temperature. There are two phases are processed in the training process. In first phase, it reads the data frame. Then it sets the list of crop which contains the data's per year, area and temperature. In second phase, it applied support vector matrices to predict the state of crop dataset.



ANN

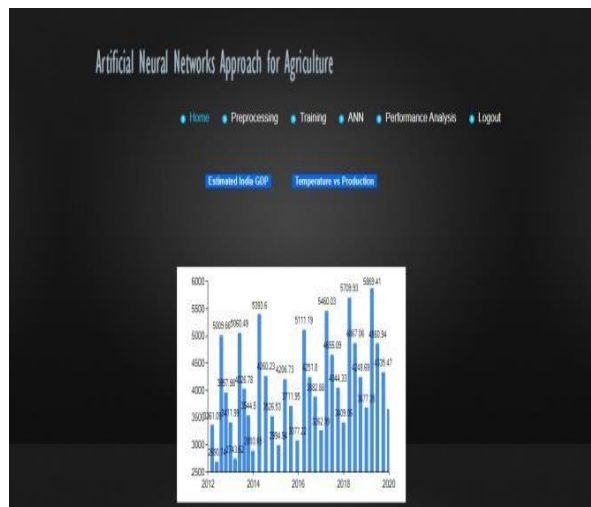
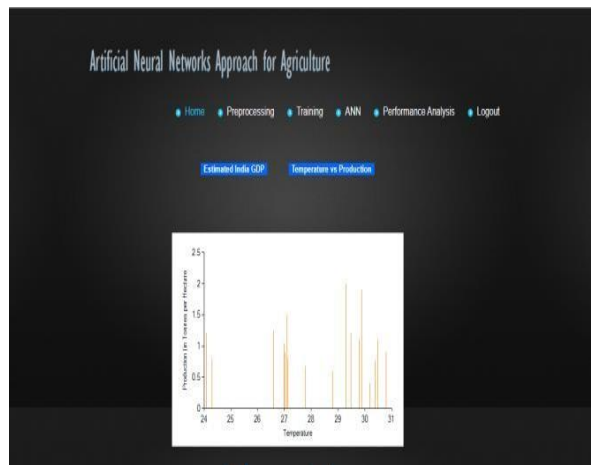
ANN algorithm used to predict the crop for year of production. In the first phase, data frame are read and

processed. Then it set the list of crop per season, temperature and by rainfall. ANN matrices to predict state for crop dataset are applied in the second phase. The regression format dataset calculated by ANN and gives result.

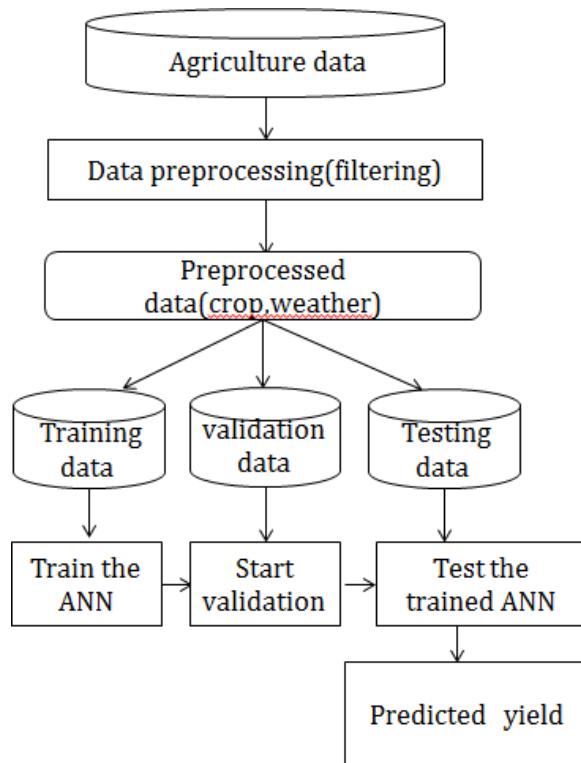


Performance Analysis:

Finally, generate graph for the output. It could understand by the high dimensional between yearly complex and seasonal climate patterns. It is to be able to predict the effect of drought and other climate conditions.



V. ARCHITECTURE DIAGRAM:



VI. CONCLUSION

In this analysis, we used some of the common data mining techniques in the field of agriculture. Some of these techniques such as the k-means, k nearest neighbor, SVM, and Bayesian network are discussed and an application in agriculture for each of these techniques is presented. Data mining in agriculture is an upcoming research field. The regression format dataset calculated by ANN and gives result.

REFERENCES

- [1] Veenadhari S, Misra B, Singh CD. Data mining techniques for predicting crop productivity—A review article. In: IJCST. 2011; 2(1).
- [2] Jain A, Murty MN, Flynn PJ. Data clustering: a review. ACM ComputSurv. 1999;31(3):264–323.
- [3] Berkhin P. A survey of clustering data mining technique. In: Kogan J, Nicholas C, Teboulle M, editors. Grouping multidimensional data. Berlin: Springer; 2006. p. 25–72
- [4] V. Arulkumar. "An Intelligent Technique for Uniquely Recognising Face and Finger Image Using Learning Vector Quantisation (LVQ)- based Template Key.