

CROP YIELD PREDICTION USING ML TECHNIQUES

Nagashree.A, Mehashree A.C
PG Scholar, Assistant Professor,
Visvesvaraya Technological University, Centre for Post graduate Studies,
Mysuru, INDIA

INTRODUCTION

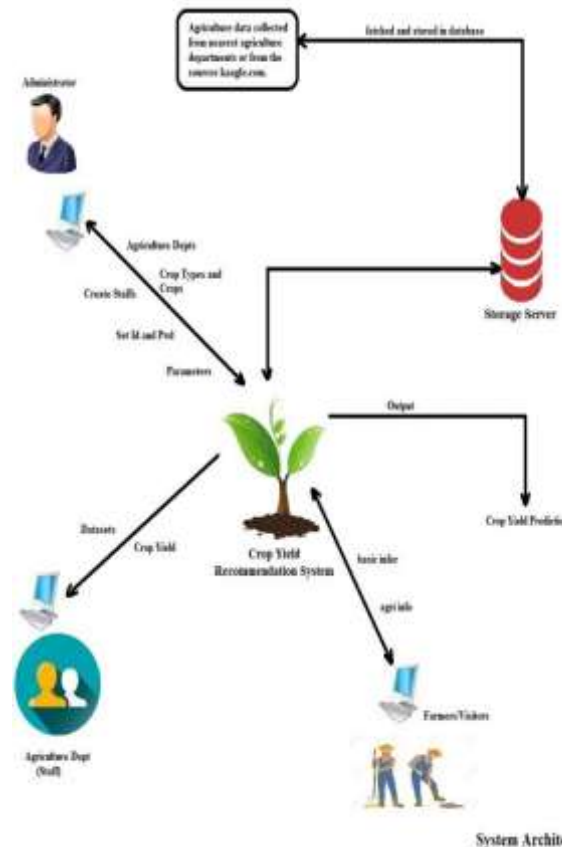
ABSTRACT

Current agriculture sector is facing so many challenges such as ways to increase the crop production, maintaining the proper agriculture parameters, maintaining proper fertilizers, reducing number of farmers suicides, increasing farmers profits etc... Food is perhaps the most important aspect of life, for without it, human kind can't survive. Prediction of crop yield helps in managing the storage of crops as well as it directs the transportation decision and risk management issues related to crops. Hence, storing food becomes an absolute compulsion for countries all over the world, especially the developing ones. The critical aspect the governments of such countries want to fulfill is storing enough food for the long term, particularly at a time of natural disaster. Hence the estimation of food is very demanding and required. In this work, we attempt to forecast the yield of few crops using data science algorithms.

India is a country where agriculture and agriculture related industries are the major source of living for the people. Agriculture is a major source of economy of the country. It is also one of the country which suffer from major natural calamities like drought or flood which damages the crop. This leads to huge financial loss for the farmers thus leading to the suicide. Predicting the crop yield well in advance prior to its harvest can help the farmers and Government organizations to make appropriate planning like storing, selling, fixing minimum support price, importing/exporting etc.

Current agriculture field facing many problems and one of the major problem that the farmers are facing is lack of profit level. Farmers grow the crops but not getting the proper yield which leads to less profit level. Yield prediction plays vital role in the agriculture department and Yield depends on many constraints such as temperature, rainfall, soil features etc. It is important to predict the yield early based on these constraints. So this leads to the development of “Yield Prediction Software”. ML can be used

vigorously to help farmers yield the best possible amount of yield.



Machine Learning

ML concerns with construction and study of system that can learn from data. For example, ML can be used in E-mail message to learn how to distinguish between spam and inbox messages. A computer program is said to be learn from experience E with respect to some task T and some performance P only if the program performance increases with experience E. ML is a branch of AI which contains statistical, probabilistic, optimization technique that can learn from past experience and discover the pattern from large complex data sets.

Types of ML:

There are three types of Machine learning(ML), they are

i. Supervised Machine Learning

Supervised learning is the types of machine learning in which machines are trained using well "labeled" training data, and on basis of that data, machines predict the output. The labeled data means some input data is already tagged with the correct output.

ii. Unsupervised Machine Learning

Unsupervised learning is a type of [algorithm](#) that learns patterns from untagged data.

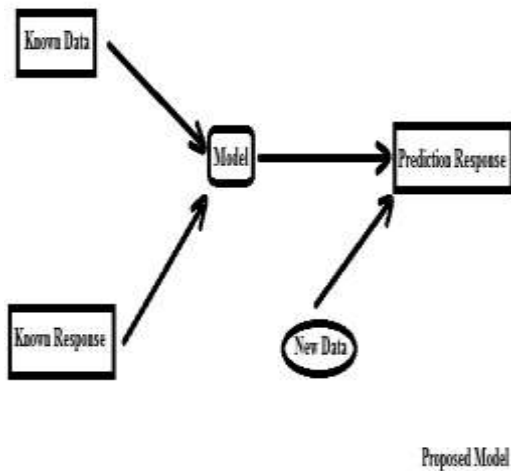
iii. Semi-Supervised Machine Learning

Mixture of both Supervised and Unsupervised Machine Learning

ML Techniques Description

Supervised Learning Technique is a predictive model used for the tasks where it involves prediction of one value using other values in the data-set. Supervised learning will have predefined labels. It classifies an object based on the parameters to one of the predefined set of labels. We have many algorithms to build model in supervised learning such as KNN, Naive bayes, Decision Tree, ID3, Random Forest, SVM , Regression techniques etc.... Depending of the requirement, labels, parameters and data-set we select the appropriate algorithm for predictions. Algorithm is

used to build a model that makes predictions based on evidence in the presence of uncertainty.



Advantages of using ML

- ❖ Continuous Improvement. Machine Learning algorithms are capable of learning from the data we provide (agriculture data).
- ❖ Automation for everything
- ❖ Less time taken to process data.
- ❖ More accurate results
- ❖ Faster decision making.
- ❖ Trends and patterns identification
- ❖ Wide range of applications

RESEARCH PROGRESS

Literature Survey

1. IEEE PAPER TITLE: Prediction of Crop Cultivation

YEAR OF PUBLICATION: 2019

AUTHORS: Neha Rale, Raxitkumar Solanki, Doina Bein, James Andro-Vasko, Wolfgang Bein.

METHODOLOGY: linear regression with polynomial features, and support-vector regression using a Radial Basis Function (RBF) kernel.

LIMITATIONS:

- Linear regression and support vector regression generates outputs graphically which is difficult to analyze.
- Not suitable in real time.

- Small Data-set used for prediction.

2. IEEE PAPER TITLE: Crop Yield Prediction and Efficient use of Fertilizers.

YEAR OF PUBLICATION: 2019

AUTHORS: S.Bhanumathi, M.Vineeth and N.Rohit.

METHODOLOGY: Random Forest and Back propagation algorithm used for implementation.

LIMITATIONS:

- Less parameters used for yield prediction
- Based on fertilizers , system predicts crop yield , but not considering all agriculture parameters.

Proposed Work

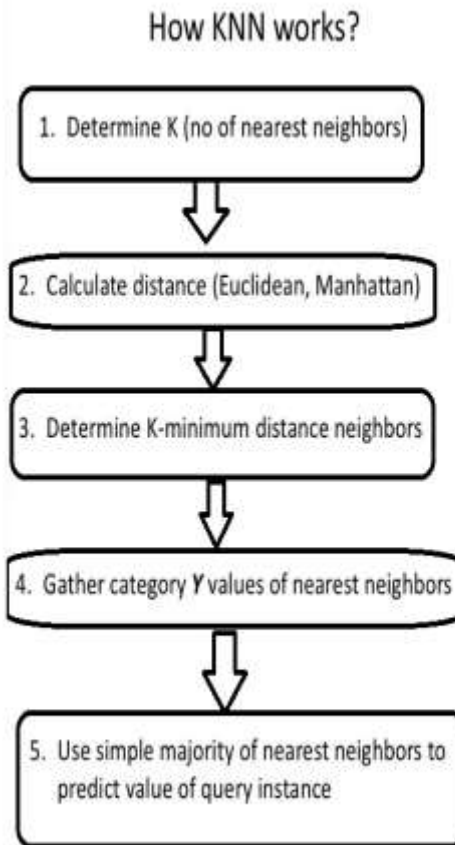
Proposed system helps farmers to cultivate right type of crops in right time and also helps farmers to increase crop yield by suggesting suitable profits. Major population in India will benefit from this application. Supervised learning algorithms used for the recommendations such as either "bayesian classifier" or "K nearest neighbor" and "Naive Bayes Algorithm". These algorithms are preferred as they work efficiently, generated faster results and also work for all formats of data. and also few survey papers suggests these algorithms are efficient and good for agriculture datasets.

Methodology Used (How ML is used to predict crop yield)

KNN algorithm

K nearest neighbor algorithm is also one of the efficient algorithm which works only for numbers. It works for numerical data. And process data faster compared to other algorithms. It works based on the distance calculation.

Flow of the Algorithm



FLOW OF KNN ALGORITHM

Sample Data-set using KNN technique

Sample Training Datasets

KNN Method

Predict YES or NO
Predict - 0 or 1
Predict Fraud or Genuine

Records	P1	P2	P3	Result
1	7	2	1	+
2	3	5	2	-
3	2	6	2	+
4	6	3	1	+
5	4	4	2	-
6	7	2	1	-
7	2	3	4	+
8	4	3	3	?

1.Determine K

1. Determine value of K

Suppose K = 3

2. Calculate distance

Coordinates of query instance are (4,3,3)

Coordinates of training instance(1) are (7,2,1)

$$D = \text{SQRT} ((7-4)^2+(2-3)^2+(1-3)^2) = 3.74165$$

3.Determine K minimum distance

K = 3

Records	P1	P2	P3	Result	distance
1	7	2	1	+	3.741657
2	3	5	2	-	2.44949 (3)
3	2	6	2	+	3.741657
4	6	3	1	+	2.828427
5	4	4	2	-	1.414214 (1)
6	7	2	1	-	3.741657
7	2	3	4	+	2.236068 (2)
8	4	3	3	?	

2.Calculate Distance

Records	P1	P2	P3	Result	distance
1	7	2	1	+	3.741657
2	3	5	2	-	2.44949
3	2	6	2	+	3.741657
4	6	3	1	+	2.828427
5	4	4	2	-	1.414214
6	7	2	1	-	3.741657
7	2	3	4	+	2.236068
8	4	3	3	?	

4. Gather category Y values of nearest neighbors

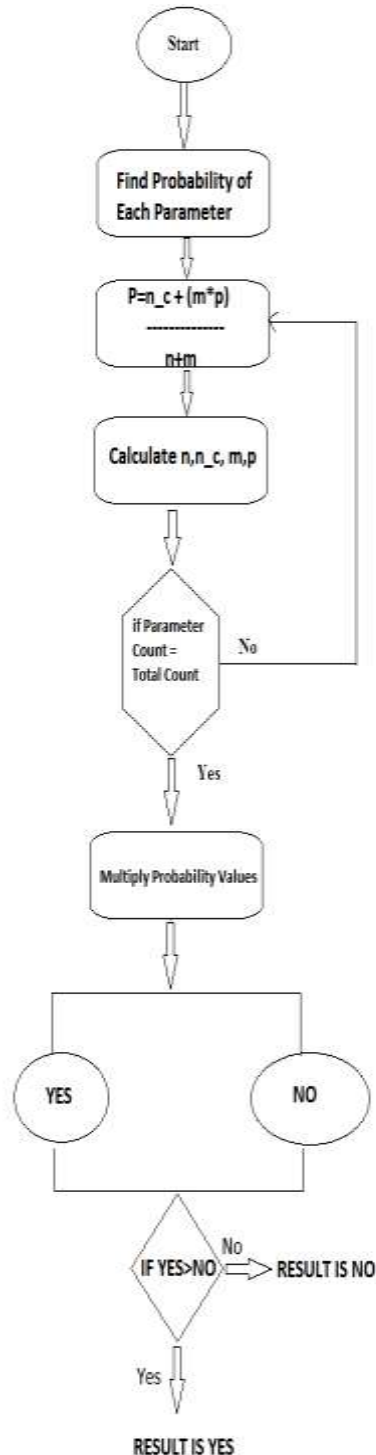
Records	P1	P2	P3	Result
2	3	5	2	-
5	4	4	2	-
7	2	3	4	+
8	4	3	3	

5. Use simple majority of nearest neighbors to predict values of query instance

Records	P1	P2	P3	Result
1	7	2	1	+
2	3	5	2	-
3	2	6	2	+
4	6	3	1	+
5	4	4	2	-
6	7	2	1	-
7	2	3	4	+
8	4	3	3	(-)

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Flow of the Algorithm



FLOW OF NAIVE BAYES ALGORITHM

Naive Bayes Algorithm

- Naïve Bayes algorithm is a supervised learning algorithm, which is based on **Bayes theorem** and used for solving classification problems.
- It is mainly used in *text classification* that includes a high-dimensional training dataset.
- Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.
- It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.
- Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles.
- The general formula for Bayes' theorem is given as:

Experiment Results

Description:

We build a model to predict the crop yield. System predicts yield for 2 different crops paddy and ragi. We use machine learning algorithm to predict yield with more accuracy and efficient results. We use training datasets to predict the crop yield. We divide the ratio 80:20 training datasets and testing datasets. For this ratio we got 96% accuracy results for paddy crop and 84% accurate results for ragi.

System took 2346 milli seconds for paddy yield prediction and 594 milli seconds for ragi yield prediction and precision factor for paddy is 96% and for wheat it is 84% and recall factor is 4% for paddy crop and 16% for ragi crop.

KNN Algorithm Results

Paddy Crop Yield

Constraint	Algorithm
Accuracy	92%
Time (milli secs)	2346
Correctly Classified (precision)	96%
InCorrectly Classified (Recall)	4%

Ragi Crop Yield

Constraint	Algorithm
Accuracy	84%
Time (milli secs)	594
Correctly Classified (precision)	84%
InCorrectly Classified (Recall)	16%

Naive Bayes Algorithm Results

Paddy Crop Yield

Constraint	Algorithm
Accuracy	96%
Time (milli secs)	2318
Correctly Classified (precision)	96%
InCorrectly Classified (Recall)	4%

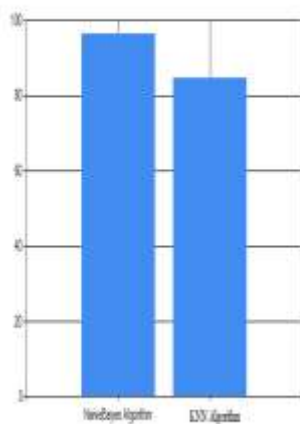
Ragi Crop Yield

Constraint	Algorithm
Accuracy	98%
Time (milli secs)	606
Correctly Classified (precision)	84%
InCorrectly Classified (Recall)	16%

Graphical Outputs

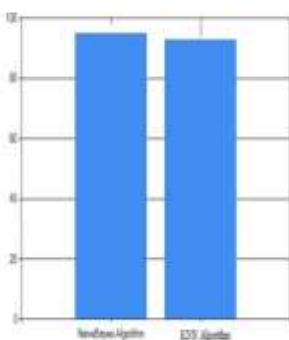
Ragi Results

Graph Representation (Algorithm Vs Accuracy)!!!



Paddy Outputs

Graph Representation (Algorithm Vs Accuracy)!!!



Conclusion

India is a country where agriculture and agriculture related industries are the major source of living for the people. Agriculture is a major source of economy of the country. India is also one of the country which suffer from major natural calamities like drought or flood which damages the crop. This leads to huge financial loss for the farmers thus leading to the suicide. Current technology “Data Science” used for crop yield prediction.

System useful for the society as it is a real time application for crop yield prediction. System helps farmers to make right decision in right time. System helps farmers to get high profits.

References

[1] Maria Rossana C.de Leon,Eugene Rex L.Jalao, “A prediction model framework for crop yield prediction,” Asia Pacific Industrial Engineering and Management System.

[2] Ramesh A. Medar,Vijay S. Rajpurohit, “A survey on Data Mining Techniques for crop yield prediction”, International Journal of Advance Research in Computer Science and Management Studies.Volume 2 , Issue 9, Sept 2014.

[3] D Ramesh, B Vishnu Vardhan, “Region specific crop yield Analysis :A data Mining approach “,UACEE International Journal of Advances in computer Science and its Applications-IJCSIA volume 3:issue 2.

[4] S.Veenadhari, Dr Bharat Mishra, Dr C D Singh, “Soyabean productivity modeling using Decision Tree Algorithms,” International Journal of Computer Applications Vol 27-No7 Aug 2011.

[5] George RuB, “Data Mining of Agricultural Yield Data:A comparison of Regression models.

- [6] Wim G.M. Bastiaanssen, Samia Ali, “A new crop yield forecasting model based on satellite measurements applied across the Indus Basin, Pakistan,” *Agriculture, Ecosystems and Environment* 94 (2003) 321–340.
- [7] Sudhanshu Sekhar Panda, Daniel P. Ames, and Suranjan Panigrahi, “Application of Vegetation Indices for Agricultural Crop Yield Prediction Using Neural Network Techniques”, *Remote Sensing* 2010, 2, 673-696; doi:10.3390/rs2030673
- [8] S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh, “Machine Learning Approach for forecasting crop yield based on climatic parameters”, 978-1-4799-2352-6/14/\$31.00 ©2014 IEEE
- [9] Jeferson Lobato Fernandes; Jansle Vieira Rocha, “Sugarcane yield estimates using time series analysis of spot vegetation images”, *Rubens Augusto Camargo Lamparelli, Sci. Agric. (Piracicaba, Braz.)*, v.68, n.2, p.139-146, March/April 2011
- [10] Yunous Vagh, Jitian Xiao, “Mining temperature profile data for shire-level crop yield prediction”, 978-1-4673-1487-9/12/\$31.00 ©2012 IEEE, *Proceedings of the 2012 International Conference on Machine Learning and Cybernetics*, Xian, 15-17 July, 2012
- [11] M. Gunasundari, Dr. T Arunkumar, Ms. R. Hemavathy, “CRY – An improved Crop Yield Prediction model using Bee Hive Clustering 423 Approach for Agricultural data sets”, *Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering*, February 21-22, 978-1-4673-5845-3/13/\$31.00 ©2013 IEEE

