International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

Enhancement of Crop Production using Data Mining and Machine Learning techniques

Kushal kumar B.N[1], Sonali sinha[2], Prashant Tiwari[3], Preksha J.W[4], Shalini N[5]

[1]Assistant Professor, [2-5]Student
[1-5]Department of Computer Science and Engineering
[1-5]KS Institute of Technology, Bengaluru - 560062, Karnataka, India

---

ABSTRACT: Currently India is facing a huge population crisis, the current population is 133.92 crore and there are predictions that India will be adding 273 million more by 2050, which will eventually make the population of the country 161 crores. Providing food security to every individual will become the challenging task and it will increase a lot more pressure on farm lands to grow more crops. But the agriculture land is decreasing at a rate of 30,000 hectares per year. Which means farmers are giving up on their agricultural land and the primary reason behind this is significant drop in the yield eventually translating into low revenue. The poor yield can be attributed to many factors such as climate change, change in the nature of soil and farmers not well equipped with information to adjust their methods to changing conditions. There is a need to make a platform which can help our farmers to take right decision on field and not to confuse them with lots of information.

The proposed model is aimed at equipping the farmers with the best, condensed and actionable knowledge that is specific to their farms and climate type. The knowledge is currently distributed all over the internet, we aspire to not only bring them in condensed form at a single platform but also make it easy for the farmers to understand the knowledge provided and translate it into actions.

In the first phase we aim to develop a combination of models to give the farmers some statistical suggestions after taking basic information from the farmers through the web interface, further we plan to develop image processing and video analysing model and integrate to web interface.

In this paper we discuss the implementation of the combination of model. The application of machine learning and data mining methodologies to give statistical suggestion is discussed. For further enhancements this model can be integrated with NLP models to give farmers more flexibility and the IOT devices to get more accurate information at micro level.

Keywords: Machine learning, data mining, and image processing, video analyzer.

1. INTRODUCTION:

Population crisis and farmers giving up on their land is directly posing a threat to the food security of every individual. Whereas agriculture is also the major source of Indian economy[1].But due to climate change there is a lot of shift in the ingredients which plays a vital role in crop production, like soil texture, water quality, seed quality, air temperature etc.

Farmers being unaware of these basic need of the crop or being overwhelmed with lots of information over the net end up getting confused. Resulting in reduced crop production, low revenue and poor living conditions compelling farmers to give up on their land. To overcome all these problems and enhance the crop yield farmers, agriculture scientists and researchers are trying new ways and technologies.

Analyzing and finding the hidden patterns from a large dataset according to various perspectives for classification and converting it into the relevant information is data mining. Some of the data mining techniques used in this project are Association rules, K-means (clustering) and neighbours, SVM (Support vector machine).

Learning from the data provided and giving predictions for the future event on the basis of statistical calculations to give insight to the relevant facts and figures is machine learning. The machine learning techniques used in this project are Convolutional Neural Network (CNN), Bayesian Network, Backpropagation and feedforward propagation.

2. DATA MINING AND MACHINE LEARNING METHODOLOGIES

2.1 Association rules

Association rule mining is finding interesting relationships or association among the large sets of data items. It basically shows how frequent the itemset occurs in a transaction.eg sales analysis. Associations rules are nothing but a if-then statements that can support to show the probability of relationships between data items within large data sets in various databases.

At basic level association rule use ML models to analyze data for patterns or co- occurrence in database. It identifies frequent if-then association calling association rule.

An association rule has two parts (1) an antecedent (if) and a consequent (then).

An antecedent is an item found within the data sets and consequent is nothing but an item found in combination with the antecedents. These rules are created by searching large data sets for frequent if-then patterns and then it use the criteria of support and confidence to find the most important relationships. Support indicates how frequently the items appear in the data sets . Confidence gives the indication of the number of times the following if-then statements are found true.

2.2 K-MEANS CLUSTERING AND K NEAREST NEIGHBOURS

2.2.1 K-means clustering

K-Means Clustering is one of the clustering methods that process a group of data points into a small number of clusters. For example, the items in a shop the clothes are clustered in different categories (medium, large, XL are grouped as the size of the dress). This is a qualitative method of splitting a group of data. A quantitative approach is used to measure unique characteristics of the products. In k clusters, the number of data points have to be partitioned. The aim of this methodology is to assign a cluster to each data point. K-means algorithm aims to find out the clusters positions that minimize the distance from the data points to the cluster.

2.2.2 K-nearest neighbours

K-Nearest Neighbors is one of the classification algorithms in Machine Learning. It is otherwise called as supervised learning model and lazy algorithm because of instance learning. It is used for different applications like pattern recognition, data mining and intrusion detection. Implementation is simple for tiny data sets. The training data does not need any knowledge about the structure of data before the analysis. Drawback of this classifier is finding out the nearest neighbour for each sample.Lot of space is needed when the training data is large. The distance between test data and the training data should be calculated for every test data. So, the testing needs a lot of time. There are two phases in this classifier: Training phase: Save the examples, Prediction phase:

Get the test instance and find the training set.

2.3 SUPPORT VECTOR MACHINES

It is one of the supervised learning which uses a hyperplane to categorize new examples. It is used for classification and regression techniques. In a given a set of training examples, each example is marked as belonging to one or the other categories. Support Vector Machine training algorithm to develop a model which assigns new set of examples to one category or the other categories, making the model to be non-probabilistic binary linear classifier. The SVM model is also a representation of all examples existing as points in space, they are mapped so that the examples of the separate categories get divided by a clear gap which is as wide as possible. New set of examples are then mapped into the same space and predict whether it belongs to a category based on the side of the gap on which they fall. Support vector machines (algorithm) can perform linear classification and non-linear classification using a trick called kernel trick, which implicitly map their inputs into high-dimensional feature spaces.

2.4 CONVOLUTIONAL NEURAL NETWORK

CNN is class of deep neural networks consisting of various hidden layers for computing the given input in various different aspects. We can also say that CNN are an upgraded versions of multilayer perceptrons which means a fully connected network where each neuron of one layer is connected to every other neuron in the next layer. CNN are mostly inspired from the biological process of neurons. It uses a mathematical operation called convolution which is a specialized linear operation. CNN is simply a neural network that convolution in place of general matrix multiplication in at least on their layer.

2.5 BAYESIAN NEURAL NETWORK.

A Bayesian network is an acyclic graph consists of edges and nodes with directions in which each edge represents a conditional dependency, and each node represents a unique random variable. The probabilistic graphical model that uses Bayesian inference for computations are called as Bayesian networks. The idea of such design is to combine the strengths of neural networks and stochastic modeling. Neural Networks exhibit universal continuous function approximator capabilities. Probabilistic models(a.k.a statistical model) allows the direct specification of a model which uses known interaction between parameters to generate data. In prediction phase, probabilistic models generate a complete posterior distribution and produce guarantees which are probabilistic in nature on the predictions. BNNs are a unique combination of neural networks and probabilistic models with the stochastic model forming the core of this integration. BNNs can then produce guarantees which are probabilistic in nature on it's predictions and generate the distribution of the parameters which it has learnt from the following observations. Which means, in the parameter space, one can deduce the nature and distribution of the neural network's learnt parameters. The above described two characteristics make them highly attractive to theoreticians as well as practitioners.

3. CONCLUSION

The study described in this paper proposes a combination of model which aims at improving the farm production by using the techniques of data mining and machine learning and not to confuse farmers with lots of unnecessary information

4. REFERENCES

[1] "**Agriculture Analysis Using Data Mining And Machine Learning Techniques**","Vanitha CN1", "Archana N2"," Sowmiya R3"(2019)

[2] "Soil Analysis and Crop Fertility Prediction", "Komal Abhang", "Surabhi Chaughule"," Pranali Chavan", "Shraddha Ganjave".(2018)

[3] "Soil Analysis and Prediction of Suitable Crop for Agriculture using Machine Learning","S. Panchamurthi. M. E"," M. D. Perarulalan"," A. Syed Hameeduddin", " P. Yuvaraj"**.**

[4] "Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach","Monali Paul"," Santosh K. Vishwakarma"," Ashok Verma".

[5] "Machine learning approach for forecasting crop yield based on climatic parameters","S.Veenadhari"," Dr. Bharat Misra"," Dr. CD Singh".

[6] "ANALYSIS OF CROP YIELD PREDICTION USING DATA MINING TECHNIQUES","D Ramesh"," B Vishnu Vardhan".