ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 8, Issue 3, pp: (253-263), Month: July - September 2020, Available at: www.researchpublish.com

PREDICTION OF CROP YIELD AND COST BY FINDING BEST ACCURACY USING MACHINE LEARNING

[1]Varadharajan V, [2]Rajasekaran S, [3]Mr. Dr. P.Mohan kumar

[1]Computer Science and Engineering, Jeppiaar SRR Engineering College, Padur, Chennai

[2]Computer Science and Engineering, Jeppiaar SRR Engineering College, Padur, Chennai

[3]Head of Department, Professor, Computer Science and Engineering, Jeppiaar SRR Engineering College, Padur, Chennai

Abstract: **Among worldwide, agriculture has the major responsibility for improving the economic contribution of the nation. However, still the most agricultural fields are under developed due to the lack of deployment of ecosystem control technologies. Due to these problems, the crop production is not improved which affects the agriculture economy. Hence a development of agricultural productivity is enhanced based on the plant yield prediction. To prevent this problem, Agricultural sectors have to predict the crop from given dataset using machine learning techniques. The analysis of dataset by supervised machine learning technique(SMLT) to capture several information's like, variable identification, uni-variate analysis, bi-variate and multi-variate analysis, missing value treatments etc. A comparative study between machine learning algorithms had been carried out in order to determine which algorithm is the most accurate in predicting the best crop. The results show that the effectiveness of the proposed machine learning algorithm technique can be compared with best accuracy with entropy calculation, precision, Recall, F1 Score, Sensitivityand Specificity.**

Keywords: **dataset, Machine learning-Classification method.**

I. INTRODUCTION

In developing countries, farming is considered as the major source of revenue for many people. In modern years, the agricultural growth is engaged by several innovations, environments, techniques and civilizations. In addition, the utilization of information technology may change the condition of decision making and thus farmers may yield the best way. For decision making process, data mining techniques related to the agriculture are used. Data mining is a process of extracting the most significant and useful information from the huge amount of datasets. Nowadays, we used machine learning approach with developed in crop or plant yield prediction since agriculture has different data like soil data, crop data, and weather data. Plant growth prediction is proposed for monitoring the plant yield effectively through the machine learning techniques.

RELATED WORKS:

1) **ZHENG Guanghui1, Dongryeol RYU2,∗, JIAO Caixia1 and HONG Changqiao1**: Estimation of Organic Matter Content in Coastal Soil Using Reflectance Spectroscopy Research, 2015.

Rapid determination of soil organic matter (SOM) using regression models based on soil reflectance spectral data serves an important function in precision agriculture. "Deviation of arch" (DOA)-based regression and partial least squares regression (PLSR) are two modeling approaches to predict SOM. However, few studies have explored the accuracy of the DOA-based regression and PLSR models.

2) **Zhiqiang Cheng 1,2 ID , JihuaMeng 1,*, YanyouQiao 1, Yiming Wang 1,2, Wenquan Dong 1 and Yanxin Han 1,2**, Preliminary Study of Soil Available Nutrient Simulation Using a Modified WOFOST(Model and Time-Series Remote Sensing Observations), 2017.

The approach of using multispectral remote sensing (RS) to estimate soil available nutrients (SANs) has been recently developed and shows promising results. This method overcomes the limitations of commonly used methods by building a statistical model that connects RS-based crop growth and nutrient content. However, the stability and accuracy of this model require improvement.

3) **Ming Jin, Xiangnan Liu, Ling Wu, and Meiling Liu**, Distinguishing Heavy-Metal Stress Levels in Rice Using Synthetic Spectral Index(Responses to Physiological Function Variations) , 2016.

Accurately assessing the heavy-metal contamination in crops is crucial to food security. This study provides a method to distinguish heavy-metal stress levels in rice using the vari- ations of two physiological functions as discrimination indices, which are obtained by assimilation of remotely sensed data with a crop growth model. Two stress indices, which correspond to dailytotalCO2 assimilationanddry-matterconversioncoefficient were incorporated into the World Food Study (WOFOST) crop growth model and calculated by assimilating the model with leaf area index (LAI), which was derived from time-series HJ1-CCD data.

4) **ManashProtimGoswami ,BabakMontazer, and UtpalSarma, Member, IEEE**, Design and Characterization of a Fringing Field Capacitive Soil Moisture Sensor, 2018.

The optimization and imple- mentation of a fringing field capacitive soil moisture sensor using the printed circuit board technology. It includes the analysis of a novel configuration of an interdigital sensor for measuring soil moisture with two existing configurations. The optimized designs were simulated by using a 3-D finite-element method and fabricated by using a copper clad board.

LIMITATION OF EXISTINGSYSTEM

- It present a crop/weeds classification approach based on a three-steps procedure. The first step is a robust pixel-wise segmentation (i.e., soil/plant) and image patches containing plants are extracted in the second step.
- a deep CNN for crop/weed classification is used.

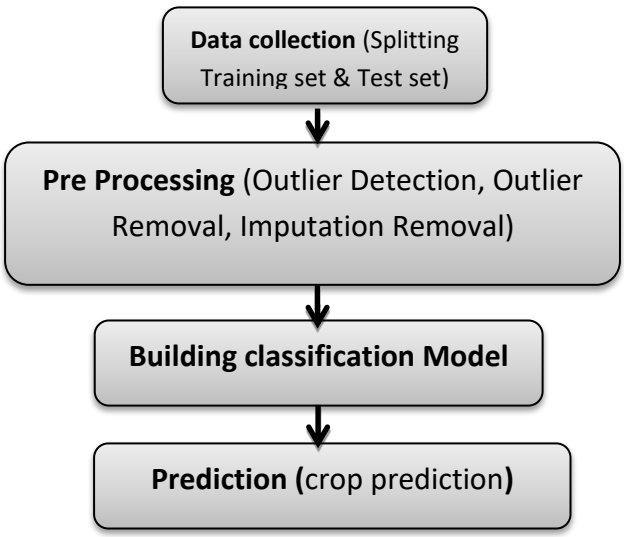
PROPOSED SYSTEM

A. In this section of the report, you will load in the data, check for cleanliness, and then trim and clean your dataset for analysis. Make sure that you document your steps carefully and justify your cleaning decisions.

➢ Our goal is push for assisting farmers, government using our predictions. All these publications state they have done better than their competitors but there is no article or public mention of their work being used practically to assist the farmers. If there are some genuine problems in rolling out that work to next stage, then identify those problems and try solving them.

➢ It is targeted to those farmers who wish to professionally manage their farm by planning, monitoring and analyzing all farming activities.

SYSTEM DESIGN



ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 8, Issue 3, pp: (253-263), Month: July - September 2020, Available at: www.researchpublish.com

MODULES

1. DATA VALIDATION AND PRE-PROCESSING TECHNIQUE
2. EXPLORATION DATA ANALYSIS OF VISUALIZATION AND TRAINING A MODEL BY GIVEN ATTRIBUTES
3. PERFORMANCE MEASUREMENTS OF LOGISTIC REGRESSION AND DECISION TREE ALGORITHMS
4. PERFORMANCE MEASUREMENTS OF SUPPORT VECTOR CLASSIFIER AND RANDOM FOREST
5. PERFORMANCE MEASUREMENTS OF KNN AND NAIVE BAYES
6. GUI BASED PREDICTION OF CROP YIELD AND YIELD COST

II. MODULE DESCRIPTIONS

1. DATA VALIDATION AND PRE-PROCESSING TECHNIQUE

data validation process:

Validation techniques in machine learning are used to get the error rate of the Machine Learning (ML) model, which can be considered as close to the true error rate of the dataset. If the data volume is large enough to be representative of the population, you may not need the validation techniques. However, in real-world scenarios, to work with samples of data that may not be a true representative of the population of given dataset. To finding the missing value, duplicate value and description of data type whether it is float variable or integer. The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyper parameters. The evaluation becomes more biased as skill on the validation dataset is incorporated into the model configuration. The validation set is used to evaluate a given model, but this is for frequent evaluation. It as machine learning engineers uses this data to fine-tune the model hyper parameters. Data collection, data analysis, and the process of addressing data content, quality, and structure can add up to a time-consuming to-do list. During the process of data identification, it helps to understand your data and its properties; this knowledge will help you choose which algorithm to use to build your model. For example, time series data can be analyzed by regression algorithms; classification algorithms can be used to analyze discrete data. (For example to show the data type format of given dataset)

	State_Name	District_Name	Crop_Year	Season	Crop	Area	rainfall	Average Humidity	Mean Temp	Cost of Cultivation (`/Hectare) C2	Cost of Production (`/Quintal) C2	Yield (Quintal/ Hectare)	cost of production per yield
0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254.0	0.012360	57	62	23076.74	1941.55	9.83	19085.4365
1	Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Arecanut	1254.0	0.084119	56	58	12610.85	1691.66	6.83	11554.0378
2	Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Arecanut	1258.0	0.080064	58	53	32683.46	3207.35	9.33	29924.5755
3	Andaman and Nicobar Islands	NICOBARS	2003	Whole Year	Arecanut	1261.0	0.181051	57	58	13209.32	2228.97	5.90	13150.9230
4	Andaman and Nicobar Islands	NICOBARS	2004	Whole Year	Arecanut	1264.7	0.035446	63	67	22560.30	1595.56	13.57	21651.7492

PREPROCESSING

Importing the library packages with loading given dataset. To analyzing the variable identification by data shape, data type and evaluating the missing values, duplicate values. A validation dataset is a sample of data held back from training your model that is used to give an estimate of model skill while tuning model's and procedures that you can use to make the best use of validation and test datasets when evaluating your models. Data cleaning / preparing by rename the given dataset and drop the column etc. to analyze the uni-variate, bi-variate and multi-variate process. The steps and techniques for data cleaning will vary from dataset to dataset. The primary goal of data cleaning is to detect and remove errors and anomalies to increase the value of data in analytics and decision making.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 8, Issue 3, pp: (253-263), Month: July - September 2020, Available at: www.researchpublish.com

```
#preprocessing, split test and dataset, split
X = df.drop(labels='CPPY', axis=1)
#Response variable
y = df.loc[:,'CPPY']
```

```
#We'll use a test size of 30%. We also stratif
from sklearn.model_selection import train_test
X_train, X_test, y_train, y_test = train_test_
print("Number of training dataset: ", len(X_tr
print("Number of testing dataset: ", len(X_tes
print("Total number of dataset: ", len(X_train
```

```
Number of training dataset:  163704
Number of testing dataset:  70160
Total number of dataset:  233864
```

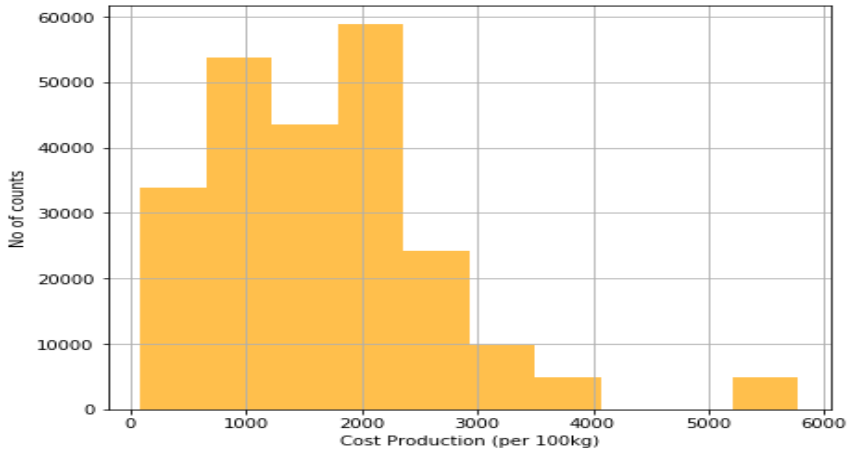
2. EXPLORATION DATA ANALYSIS OF VISUALIZATION AND TRAINING A MODEL BY GIVEN ATTRIBUTES

Exploration data analysis of visualization:

Data visualization is an important skill in applied statistics and machine learning. Statistics does indeed focus on quantitative descriptions and estimations of data. Data visualization provides an important suite of tools for gaining a qualitative understanding. This can be helpful when exploring and getting to know a dataset and can help with identifying patterns, corrupt data, outliers, and much more. With a little domain knowledge, data visualizations can be used to express and demonstrate key relationships in plots and charts that are more visceral and stakeholders than measures of association or significance. Data visualization and exploratory data analysis are whole fields themselves and it will recommend a deeper dive into some the books mentioned at the end.

Sometimes data does not make sense until it can look at in a visual form, such as with charts and plots. Being able to quickly visualize of data samples and others is an important skill both in applied statistics and in applied machine learning. It will discover the many types of plots that you will need to know when visualizing data in Python and how to use them to better understand your own data.

- How to chart time series data with line plots and categorical quantities with bar charts.
- How to summarize data distributions with histograms and box plots.
- How to summarize the relationship between variables with scatter plots.



3.PERFORMANCE MEASUREMENTS OF LOGISTIC REGRESSION AND DECISION TREE ALGORITHMS

Logistic Regression**:**

It is a statistical method for analysing a data set in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes). The goal

of logistic regression is to find the best fitting model to describe the relationship between the dichotomous characteristic of interest (dependent variable = response or outcome variable) and a set of independent (predictor or explanatory) variables. Logistic regression is a Machine Learning classification algorithm that is used to predict the probability of a categorical dependent variable. In logistic regression, the dependent variable is a binary variable that contains data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.).

Decision Tree:

It is one of the most powerful and popular algorithm. Decision-tree algorithm falls under the category of supervised learning algorithms. It works for both continuous as well as categorical output variables. Assumptions of Decision tree:

- At the beginning, we consider the whole training set as the root.
- Attributes are assumed to be categorical for information gain, attributes are assumed to be continuous.
- On the basis of attribute values records are distributed recursively.
- We use statistical methods for ordering attributes as root or internal node.

Decision tree builds classification or regression models in the form of a tree structure. It breaks down a data set into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. A decision node has two or more branches and a leaf node represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data. Decision tree builds classification or regression models in the form of a tree structure. It utilizes an if-then rule set which is mutually exclusive and exhaustive for classification. The rules are learned sequentially using the training data one at a time. Each time a rule is learned, the tuples covered by the rules are removed.

4. PERFORMANCE MEASUREMENTS OF SUPPORT VECTOR CLASSIFIER AND RANDOM FOREST

Support Vector Machines (SVM):

A classifier that categorizes the data set by setting an optimal hyper plane between data. I chose this classifier as it is incredibly versatile in the number of different kernelling functions that can be applied and this model can yield a high predictability rate. Support Vector Machines are perhaps one of the most popular and talked about machine learning algorithms. They were extremely popular around the time they were developed in the 1990s and continue to be the go-to method for a high-performing algorithm with little tuning.

- How to disentangle the many names used to refer to support vector machines.
- The representation used by SVM when the model is actually stored on disk.
- How a learned SVM model representation can be used to make predictions for new data.
- How to learn an SVM model from training data.
- How to best prepare your data for the SVM algorithm.
- Where you might look to get more information on SVM.

Random Forest:

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of over fitting to their training set. Random forest is a type of supervised machine learning algorithm based on ensemble learning. Ensemble learning is a type of learning where you join different types of algorithms or same algorithm multiple times to form a more powerful prediction model. The random forest algorithm combines multiple algorithm of the same type i.e. multiple decision *trees*, resulting in a *forest of trees*, hence the name "Random Forest". The random forest algorithm can be used for both regression and classification tasks.

The following are the basic steps involved in performing the random forest algorithm:

- Pick N random records from the dataset.
- Build a decision tree based on these N records.
- Choose the number of trees you want in your algorithm and repeat steps 1 and 2.
- In case of a regression problem, for a new record, each tree in the forest predicts a value for Y (output). The final value can be calculated by taking the average of all the values predicted by all the trees in forest. Or, in case of a classification problem, each tree in the forest predicts the category to which the new record belongs. Finally, the new record is assigned to the category that wins the majority vote.

5. PERFORMANCE MEASUREMENTS OF KNN AND NAIVE BAYES

K-Nearest Neighbor (KNN):

K-Nearest Neighbor is a supervised machine learning algorithm which stores all instances correspond to training data points in n-dimensional space. When an unknown discrete data is received, it analyzes the closest k number of instances saved (nearest neighbors) and returns the most common class as the prediction and for real-valued data it returns the mean of k nearest neighbors. In the distance-weighted nearest neighbor algorithm, it weights the contribution of each of the k neighbors according to their distance using the following query giving greater weight to the closest neighbors.

Usually KNN is robust to noisy data since it is averaging the k-nearest neighbors. The k-nearest-neighbors algorithm is a classification algorithm, and it is supervised: it takes a bunch of labeled points and uses them to learn how to label other points. To label a new point, it looks at the labeled points closest to that new point (those are its nearest neighbors), and has those neighbors vote, so whichever label the most of the neighbors have is the label for the new point (the "k" is the number of neighbors it checks). Makes predictions about the validation set using the entire training set. KNN makes a prediction about a new instance by searching through the entire set to find the k "closest" instances. "Closeness" is determined using a proximity measurement (Euclidean) across all features.

Naive Bayes algorithm:

The Naive Bayes algorithm is an intuitive method that uses the probabilities of each attribute belonging to each class to make a prediction. It is the supervised learning approach you would come up with if you wanted to model a predictive modeling problem probabilistically. Naive bayes simplifies the calculation of probabilities by assuming that the probability of each attribute belonging to a given class value is independent of all other attributes. This is a strong assumption but results in a fast and effective method. The probability of a class value given a value of an attribute is called the conditional probability. By multiplying the conditional probabilities together for each attribute for a given class value, we have a probability of a data instance belonging to that class. To make a prediction we can calculate probabilities of the instance belonging to each class and select the class value with the highest probability.

Naive Bayes is a statistical classification technique based on Bayes Theorem. It is one of the simplest supervised learning algorithms. Naive Bayes classifier is the fast, accurate and reliable algorithm. Naive Bayes classifiers have high accuracy and speed on large datasets. Naive Bayes classifier assumes that the effect of a particular feature in a class is independent of other features. For example, a loan applicant is desirable or not depending on his/her income, previous loan and transaction history, age, and location. Even if these features are interdependent, these features are still considered independently. This assumption simplifies computation, and that's why it is considered as naive. This assumption is called class conditional independence.

6. GUI BASED PREDICTION OF CROP YIELD AND YIELD COST

Tkinter is a python library for developing GUI (Graphical User Interfaces). We use the tkinter library for creating an application of UI (User Interface), to create windows and all other graphical user interface and Tkinter will come with Python as a standard package, it can be used for security purpose of each users or accountants. There will be two kinds of pages like registration user purpose and login entry purpose of users. In the example below 4 different algorithms are compared:

- Logistic Regression
- Random Forest
- Decision tree
- Support Vector Machines

- Now, the dimensions of new features in a numpy array called 'n' and it want to predict the species of this features and to do using the predict method which takes this array as input and spits out predicted target value as output.

- So, the predicted target value comes out to be 0. Finally to find the test score which is the ratio of no. of predictions found correct and total predictions made and finding accuracy score method which basically compares the actual values of the test set with the predicted values.

Sensitivity:

Sensitivity is a measure of the proportion of actual positive cases that got predicted as positive (or true positive). Sensitivity is also termed as Recall. This implies that there will be another proportion of actual positive cases, which would get predicted incorrectly as negative (and, thus, could also be termed as the false negative). This can also be represented in the form of a false negative rate. The sum of sensitivity and false negative rate would be 1. Let's try and understand this with the model used for predicting whether a person is suffering from the disease. Sensitivity is a measure of the proportion of people suffering from the disease who got predicted correctly as the ones suffering from the disease. In other words, the person who is unhealthy actually got predicted as unhealthy.

Mathematically, sensitivity can be calculated as the following:

Sensitivity = (True Positive) / (True Positive + False Negative)

The following is the details in relation to True Positive and False Negative used in the above equation.

- True Positive = Persons predicted as suffering from the disease (or unhealthy) are actually suffering from the disease (unhealthy); In other words, the true positive represents the number of persons who are unhealthy and are predicted as unhealthy.

- False Negative = Persons who are actually suffering from the disease (or unhealthy) are actually predicted to be not suffering from the disease (healthy). In other words, the false negative represents the number of persons who are unhealthy and got predicted as healthy. Ideally, we would seek the model to have low false negatives as it might prove to be life-threatening or business threatening.

The higher value of sensitivity would mean higher value of true positive and lower value of false negative. The lower value of sensitivity would mean lower value of true positive and higher value of false negative. For healthcare and financial domain, models with high sensitivity will be desired.

Specificity:

Specificity is defined as the proportion of actual negatives, which got predicted as the negative (or true negative). This implies that there will be another proportion of actual negative, which got predicted as positive and could be termed as false positives. This proportion could also be called a false positive rate. The sum of specificity and false positive rate would always be 1. Let's try and understand this with the model used for predicting whether a person is suffering from the disease. Specificity is a measure of the proportion of people not suffering from the disease who got predicted correctly as the ones who are not suffering from the disease. In other words, the person who is healthy actually got predicted as healthy is specificity.

Mathematically, specificity can be calculated as the following:

Specificity = (True Negative) / (True Negative + False Positive)

The following is the details in relation to True Negative and False Positive used in the above equation.

- True Negative = Persons predicted as not suffering from the disease (or healthy) are actually found to be not suffering from the disease (healthy); In other words, the true negative represents the number of persons who are healthy and are predicted as healthy.

- False Positive = Persons predicted as suffering from the disease (or unhealthy) are actually found to be not suffering from the disease (healthy). In other words, the false positive represents the number of persons who are healthy and got predicted as unhealthy.

The higher value of specificity would mean higher value of true negative and lower false positive rate. The lower value of specificity would mean lower value of true negative and higher value of false positive.

Prediction result by accuracy:

Logistic regression algorithm also uses a linear equation with independent predictors to predict a value. The predicted value can be anywhere between negative infinity to positive infinity. We need the output of the algorithm to be classified variable data. Higher accuracy predicting result is logistic regression model by comparing the best accuracy.

True Positive Rate(TPR) = TP / (TP + FN)

False Positive rate(FPR) = FP / (FP + TN)

Accuracy: The Proportion of the total number of predictions that is correct otherwise overall how often the model predicts correctly defaulters and non-defaulters.

Accuracy calculation:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations. One may think that, if we have high accuracy then our model is best. Yes, accuracy is a great measure but only when you have symmetric datasets where values of false positive and false negatives are almost same.

Precision: The proportion of positive predictions that are actually correct. (When the model predicts default: how often is correct?)

Precision = TP / (TP + FP)

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. The question that this metric answer is of all passengers that labeled as survived, how many actually survived? High precision relates to the low false positive rate. We have got 0.788 precision which is pretty good.

Recall: The proportion of positive observed values correctly predicted. (The proportion of actual defaulters that the model will correctly predict)

Recall = TP / (TP + FN)

Recall (Sensitivity) - Recall is the ratio of correctly predicted positive observations to the all observations in actual class - yes.

F1 Scoreis the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. Intuitively it is not as easy to understand as accuracy, but F1 is usually more useful than accuracy, especially if you have an uneven class distribution. Accuracy works best if false positives and false negatives have similar cost. If the cost of false positives and false negatives are very different, it's better to look at both Precision and Recall.

General Formula:

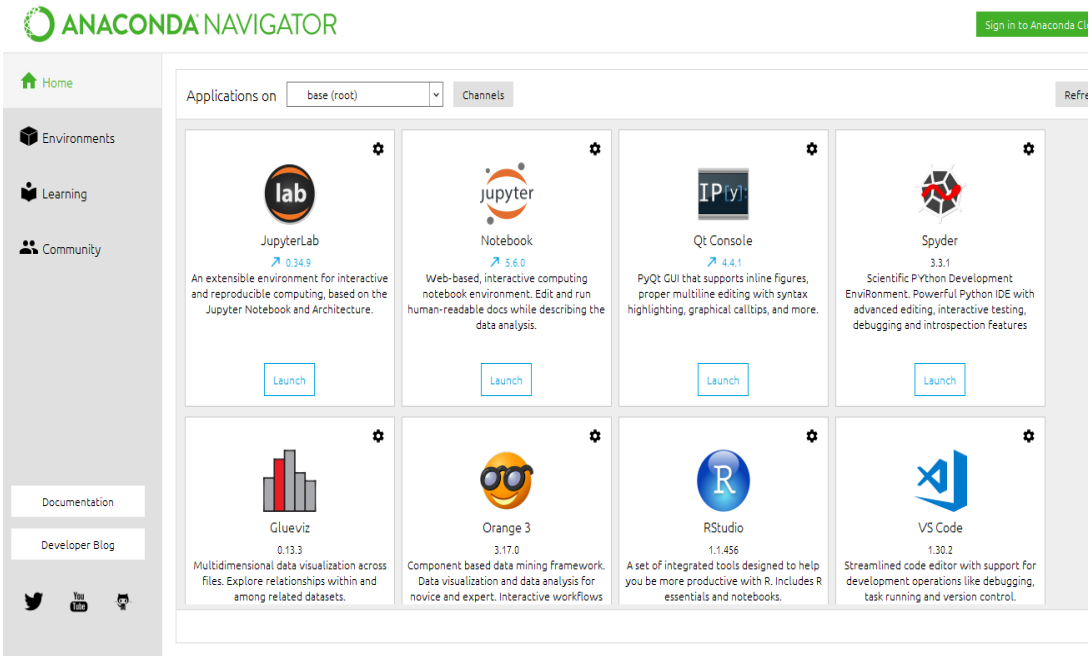
F- Measure = 2TP / (2TP + FP + FN)

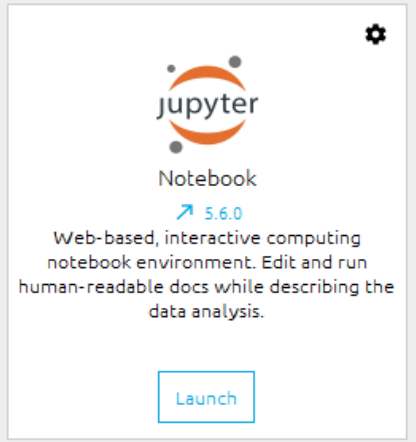
F1-Score Formula:

F1 Score = 2*(Recall * Precision) / (Recall + Precision)

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 8, Issue 3, pp: (253-263), Month: July - September 2020, Available at: www.researchpublish.com

III. RESULT AND SCREENSHOTS





IV. CONCLUSION

The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. Finally we predict the crop using machine learning algorithm with different results. This brings some of the following insights about crop prediction. As maximum types of crops will be covered under this system, farmer may get to know about the crop which may never have been cultivated and lists out all possible crops, it helps the farmer in decision making of which crop to cultivate. Also, this system takes into consideration the past production of data which will help the farmer get insight into the demand and the cost of various crops in market.

REFERENCES

[1] Addiscott, T.M., and R.J. Wagenet, 1985. Concept of solute leachingin soils: Review of modeling approaches, *Journal of Soil Science*, 36:411–424.

[2] AgRISTARS Program, 1981. *AgRISTARS: Agriculture and ResourcesInventory Surveys through Aerospace Remote Sensing*, Annual Report-Fiscal Year 1980, NASA, Lyndon B. Johnson SpaceCenter, Houston, Texas, 59 p.

[3] Allen, R., G.A. Hanuschak, and M.E. Craig, 1994. Forecasting crop acreages and yield in the face of and in spite of floods, *Crop Yield Forecasting Methods*, *Proceedings of the Seminar Villefranche-sur-Mer*, 24–27 October, Villefranche-sur-Mer, France,pp. 87–110.

[4] Asrar, G., E.T. Kanemasu, T.D. Jackson, and J.R. Pinter, 1985. Estimation of total above-ground phytomass production using remotely sensed data, *Remote Sensing of Environment*, 17:211–220.

[5] Bauer, M.E., C.T.S. Daughtry, and V.C. Vanderbilt, 1980. Spectralagronomic relationships of corn, soybean, and wheat canopies, *Proceedings of the International Colloquium on Spectral Signatures of Objects in Remote Sensing*, 08–11 September, Avignon,France, 21 p.

[6] Boatwright, G.O., and V.S. Whitefield, 1986. Early warning and crop condition assessment research, *IEEE Transactions on Geosciences and Remote Sensing*, GE-24(1):56–64.

[7] Bryant, K.J., V.W. Benson, J.R. Kiniry, J.R. Williams, and R.D. Lacewell, 1992. Simulating corn yield response to irrigation timings: Validation of the EPIC model. *Journal of Production Agriculture*, 5:237–242.

[8] Cook, P.W., R. Mueller, and P.C. Doraiswamy, 1996. Southeastern North Dakota Landsat TM crop mapping project, *Proceedings of the Annual Convention & Exhibition, ASPRS/ACSM*, 22–25 April, Baltimore, Maryland (American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland),1:600–614.

[9] Daughtry, C.S.T., K.P. Gallo, and M.E. Bauer, 1983. Spectral estimation of solar radiation intercepted by corn canopies. *Agronomy Journal*, 75:527–531.

[10] Daughtry, C.S.T., K.P. Gallo, L.L. Biehl, E.T. Kanemasu, G. Asrar, B.L. Blad, J.M. Norman, and B.R. Gardner, 1984. Spectral estimates of agronomic characteristics of crops, *Proceedings of the 10th International Symposium on Machine Processing of Remotely Sensed Data*, 12–14 April, LARS/Purdue University,West Lafayette, Indiana, pp. 348–356.

[11] Doraiswamy, P.C., T. Hodges, and D.E. Phinney, 1979. *Crop Yield Literature for AgRISTARS Crops Corn, Soybeans, Wheat, Barley, Sorghum, Rice, Cotton and Sunflowers*, AgRISTARS Technical Report SR-L9-00405, Lockheed Electronics Company, Inc., Houston, Texas, 105 p.

[12] Doraiswamy, P.C., and P.W. Cook, 1995. Spring wheat yield assessment using NOAA AVHRR data, *Canadian Journal of Remote Sensing*, 21(1):43–51.

[13] Doraiswamy, P.C., S. Hollinger, T.R. Sinclair, A. Stern, B. Akhmedov, and J. Pruger, 2001. Application of MODIS derived parameters for regional yield assessment, *Proceedings, Remote Sensing for Agriculture, Ecosystems, and Hydrology III*, 17–21 September, Toulouse, France, CD-ROM 4542–1:1–8.

[14] Edwards, D.R., V.W. Benson, J.R. Williams, T.C. Daniels, J. Lemunyon, and R.G. Gilbert, 1994. Use of the EPIC model to predict runoff transport of surface-applied inorganic fertilizer and poultry manure constituents, *Transactions of the ASAE*, 37(2):403–409.

[15] Engel, T., G. Hoogenboom, J.W. Jones, and P.W. Wilkens, 1997. AEGIS/WIN: A computer program for the application of crop simulation models across geographic areas, *Agronomy Journal*, 89(6):919–928.

[16] Gallo, K.P., C.T.S. Daughtry, and M.E. Bauer, 1985. Spectral estimators of absorbed photosynthetically active radiation in corn canopies, *Remote Sensing of Environment*, 17:221–232.

[17] Goward, S.N., D. Dye, A. Kerber, and V. Kalb, 1987. Comparison of north and south biomass from AVHRR observations, *Geocarto International*, 1:27–39.

[18] Groten, S.M.E., 1993. NDVI-crop monitoring and early yield assessment of Brukina Faso, *International Journal of Remote Sensing*, 14(8):1495–1515.

[19] Guerif, M., S. de Brisis, and B. Seguin, 1993. Combined NOAAAVHRR and SPOT-HRV data for assessing crop yields of semiarid environments, *EARsel Advances in Remote Sensing*, 2(2):110–123.

[20] Hixson, M.M., B.J. Davis, and M.E. Bauer, 1981. Sampling Landsat classifications for crop area estimation, *Photogrammetric Engineering & Remote Sensing*, 47(9):1343–1348.

[21] Jackson, R.D., and P.J. Pinter, Jr., 1986. Spectral responses of architecturally different wheat canopies, *Remote Sensing of Environment*, 20:43–56.

[22] Maas, S.J., 1988. Using satellite data to improve model estimates of crop yield, *Agronomy Journal*, 80:655–662.

[23] ———, 1993. Parameterized model of gramineous crop growth: II. Within-season simulation calibration, *Agronomy Journal*, 85:354–358.

[24] Maas, S.J., M.S. Moran, and R.D. Jackson, 1992. Combining remote sensing and modeling for regional resource monitoring, Part II: A simple model for estimating surface evaporation and biomass production, *Technical Papers, ASPRS/CSM/RT92 Convention*, 02–08 August, Washington, D.C. (American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland), pp. 225–234.

[25] Maas, S.J., M.S. Moran, A.M. Weltz, and J.H. Blanford, 1993. Model for simulating surface evaporation and biomass production using routine meteorological and remote sensing data, *Technical Papers, 1993 ACSM/ASPRS Convention*, 16–19 February,New Orleans, Louisiana (American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland), pp. 212–221.

[26] MacDonald, R.B. (editor), 1979. *The LACIE Symposium, Proceedings of Technical Sessions*, 23–26 October 1978, NASA, Lyndon B. Johnson Space Center, Houston, Texas, 1125 p.

[27] Malingreau, J.P., 1986. Global vegetation dynamics: Satellite observations over Asia, *International Journal of Remote Sensing*, 7:1121–1146.

[28] Meisinger, J.J., W.L. Hargrove, R.L. Mickelsen, J.R. Williams, and V.W. Benson, 1991. Effects of cover crops on groundwater quality, *Cover Crops for Clean Water* (H.L. Hargrove, editor), Soil and Water Conservation Society, Ankeny, Iowa, pp. 57–68.

[29] Moulin, S., A. Fisher, G. Edieu, and R. Delcolle, 1995. Temporal variation in satellite reflectances at field and regional scales compared with values simulated by linking crop growth and SAIL models, *Remote Sensing of Environment*, 54:261–272.

[30] Myers, V.L., 1983. Remote sensing applications in agriculture, *Manual of Remote Sensing, Second Edition* (R.N. Colwell, editor), American Society of Photogrammetry, Falls Church, Virginia, 2:2111–2218.

[31] NDASS, 1995. *North Dakota Agricultural Statistics, 1995*, Ag Statistics No.64, complied by the North Dakota Agricultural Statistics Service, published in cooperation with North Dakota Tanji, K.K., 1982. Modeling of the soil nitrogen cycle, *Nitrogen in Agricultural Soils* (F.J. Stevenson, editor), American Society of Agronomy, Madison, Wisconsin, pp. 721–772.