International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 05 | May 2022 www.irjet.net p-ISSN: 2395-0072

Agricultural Product Price and Crop Cultivation Prediction based on Data Science Technique

Dr. K. Jayasakthi Velmurugan[1], Divyashini D[2], Deeksha Poornima V[3]

[1]Associate Professor, Dept. of Computer Science, Jeppiaar Engineering college, Tamil Nadu, India
[2-3]Student, Dept. of Computer Science, Jeppiaar Engineering College, Tamil Nadu, India.

Abstract - *Among around the world, farming has the significant obligation regarding working on the monetary commitment of the country. In any case, still the most agricultural fields are immature because of the absence of sending of biological system control innovations. Because of these issues, the harvest creation is not further developed which influences the farming economy. Consequently an advancement of agrarian efficiency is improved in view of the plant yield expectation. To forestall this issue, Agricultural areas need to foresee the harvest from given dataset utilizing AI methods. The investigation of dataset by directed AI technique which is SMLT to catch a few data's like, variable ID, uni-variate examination, bi-variate and multi-variate investigation, missing worth medicines and so on. A near report between AI calculations had been done to figure out which calculation is the most reliable in anticipating the best harvest. The outcomes show that the viability of the proposed AI calculation procedure can measure up to best exactness with entropy computation, accuracy, Recall, F1 Score, Sensitivity, Specificity and Entropy.*

Key Words: **Supervised Machine Learning Technique, Entropy, Crop Production, Price prediction, Yield, Agriculture.**

1. INTRODUCTION

In our country, agribusiness is the primary mainstay of the economy. Most of families are reliant on agribusiness. The country's GDP is basically centered around agriculture. The greater part of the land is utilized for agribusiness to address the issues of the number of inhabitants in the district. It is important to modernize farming practices to meet the requesting necessities. Our exploration means to tackle the issue of harvest cost forecast all the more successfully to guarantee ranchers' wages. To concoct improved arrangements, it utilizes Machine Learning strategies on various information. Agri-technology and precision farming, now termed virtual farming, have emerged as new scientific areas of interest that use data-intensive methods to boost agricultural productivity and reduce the impact on the environment.

2. EXISTING SYSTEM

Crop development forecast is a vital piece of agribusiness and is essentially founded on variables, for example, soil, natural highlights like precipitation and temperature, and the quantum of compost utilized, especially nitrogen and phosphorus. These elements, be that as it may, fluctuate from one locale to another: subsequently, ranchers can't develop comparative harvests in each district. Foreseeing a reasonable yield for development is basic to agriculture**.** In this work, the MRFE is a clever methodology, has been proposed for choosing notable highlights utilizing a change crop informational collection and a positioning strategy to distinguish the most reasonable yield for a specific locale.

2.1 Drawbacks of the existing system

The primary drawbacks of the existing system are: They are not using any machine learning and deep learning concepts and they have not mentioned any metrics reports.

3. PROPOSED SYSTEM

The goal is to develop a machine learning model for Crop yield Prediction, to potentially replace the updatable supervised machine learning classification models by predicting results in the form of best accuracy by comparing supervised algorithm.

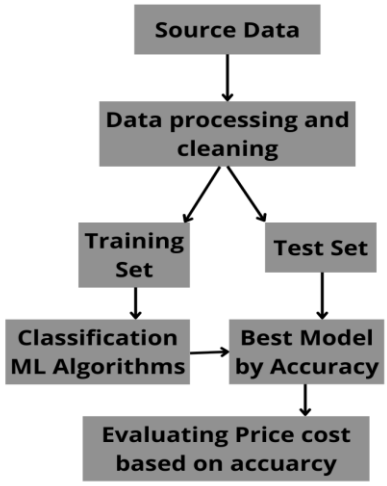


Fig -1: Workflow Diagram

3.1 Advantages

Our objective is push for helping ranchers, government utilizing our expectations. This large number of distributions state they have shown improvement over their rivals however there is no article or public notice of their work being utilized basically to help the farmers. Assuming there are a certified issues in carrying out that work to next arrange, then recognize those issues and have a go at addressing them.

3.2 Modules

Core modules of the proposed system are: Data Preprocessing, Data Analysis of Visualization, Comparing Algorithm with prediction in the form of best accuracy result and Deployment Using GUI.

4. DATASET

The demo dataset is currently provided to AI model based on this informational collection the model is prepared. Each new detail occupied at the hour of utilization structure goes about as a test informational index. After the activity of testing, model forecast in view of the surmising it finishes up based on the preparation informational indexes. Satellite Imagery (Remote Sensing Data), has been broadly utilized for anticipating crop yield. This dataset is gathered utilizing the sensors mounted on satellites or planes, which recognize the energy (electromagnetic waves), reflected or diffracted from surface of the earth. Remote detecting information has a ton of energy groups to offer, yet principally just not many of them have been utilized for crop yield expectation. However, there are certain individuals who have had a go at creating applicable highlights utilizing the groups which are commonly disregarded, and they have been fruitful with further developing outcomes with that. If there should arise an occurrence of this dataset, a great many people seldom investigate the high-request snapshots of the highlights.

Variable	Description
crop	crop name
State Name	Indian State Name
District Name	District name list of each state
Cost of Cultivation per hectare(C2	Cultivation amount for C2 Scheme
Cost of Production per Quintal(C2	Production amount for A2+FL scheme
Yeild(Quintal/Hectare)	Yield of crop
Crop Year	Crop year list
District Name	District name for each state
Area	Total area of each place
Rainfall	Water availability of each crop
Average Humidity	Directly influences the water relations of plant and indirectly affects
Mean Temprature	Climate of each crop
Cost Production per yield crop	Cost of Crop yield

Fig -2: Description of the Dataset

5. ARCHITECTURE DIAGRAM

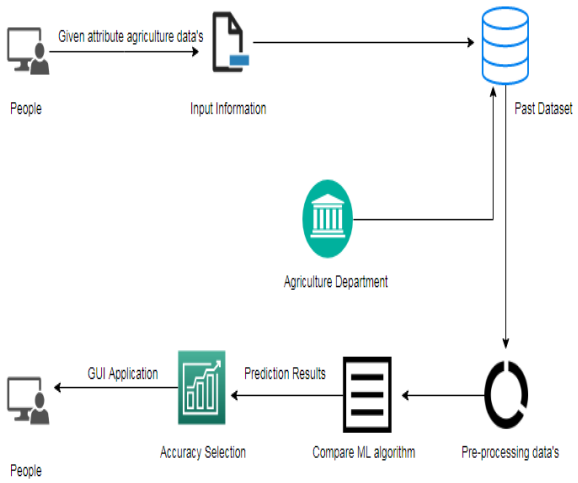


Fig -3: Architecture Diagram

6. METHODOLOGY USED

Everything the investigation of the informational index was finished utilizing —ANACONDA NAVIGATOR USING JUPYTER NOTEBOOK. It performs errands like preprocessing, order, relapse, bunching and representation. The informational index must be taken care of into the product and wanted task is chosen. It gives number of classifiers to building models and tackles insightful issues. It has the intelligent Graphical User Interface (GUI) with every one of the choices that are expected for information examination.

6.1 Algorithm Explanation

In AI and measurements, arrangement is an administered gaining approach in which the PC program gains from the information input given to it and afterward utilizes this figuring out how to group novel perception. This informational index may just be bi-class (like distinguishing whether the individual is male or female or that the mail is spam or non-spam) or it could be multi-class as well. A few instances of order issues are: discours and penmanship acknowledgment, bio metric distinguishing proof, archive grouping and so forth. In Supervised Learning, calculations gain from marked information. Subsequent to understanding the information, the calculation figures out which name ought to be given to new information in light of example and partner the examples to the unlabeled new information.

6.1.1 Decision Tree Classifier

It is one of the most remarkable and famous calculation. Choice tree calculation falls under the classification of managed learning calculations. It works for both nonstop as well as all out yield factors. Decision tree fabricates order or relapse models as a tree structure. It breaks down an informational index into increasingly small subsets while simultaneously a related choice tree is gradually evolved.

6.1.2 Random Forest Classifier

Random timberlands or arbitrary choice backwoods are a troupe learning strategy for characterization, relapse and different assignments, that work by building a huge number of choice trees at preparing time and yielding the class that is the method of the classes (arrangement) or mean forecast (relapse) of the singular trees. Irregular choice backwoods right for choice trees' propensity for over fitting to their preparation set. Random Forest is a kind of administered AI calculation in view of gathering learning.

6.1.3 Naive Bayes Algorithm

The Naive Bayes calculation is a natural technique that utilizes the probabilities of each trait having a place with each class to make an expectation. It is the managed learning approach you would concoct if you had any desire to probabilistically show a prescient displaying issue. Naïve bayes works on the computation of probabilities by expecting that the likelihood of each quality having a place with a given class esteem is free of any remaining credits.

6.1.4 Support Vector Machine Algorithm

Support-vector machines are directed learning models with related learning calculations that examine information utilized for grouping and relapse examination.

7. WORKING

7.1 Data validation Process

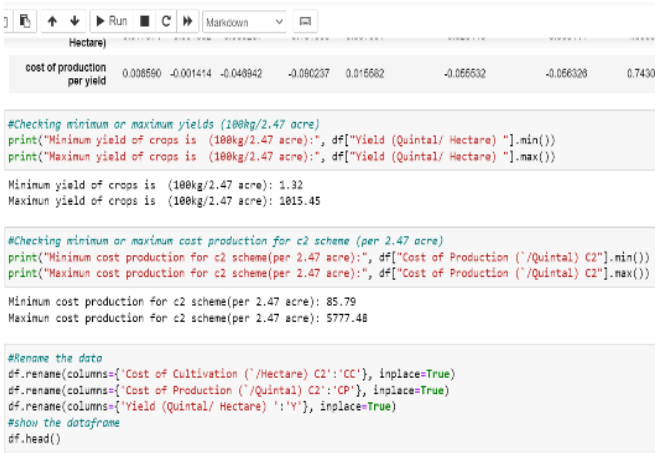
Data validation process is the blunder pace of the Machine Learning (ML) model, which can be considered as near the genuine blunder pace of the dataset. If the information volume is sufficiently enormous to be illustrative of the populace, you may not require the approval procedures. Notwithstanding, in true situations, to work with tests of information that may not be a genuine delegate of the number of inhabitants in given dataset . To viewing as the missing worth, copy worth and depiction of information type whether it is float variable or number. The example of information used to give a fair-minded assessment of a model fit on the preparing dataset while tuning model hyper boundaries. The assessment turns out to be more one-sided as expertise on the approval dataset is integrated into the model arrangement. The approval set is utilized to assess a given model, yet this is for successive assessment.

```
In [19]: #Checking for duplicate data
         df.duplicated()

Out[19]: 0       False
         1       False
         2       False
         3       False
         4       False
                 ...
         9537    False
         9538    False
         9539    False
         9540    False
         9541    False
         Length: 9542, dtype: bool

In [20]: #find sum of duplicate data
         sum(df.duplicated())
```

Fig -4: Checking Duplicate Values



```
Hectare)
cost of production per yield  0.008590  -0.001414  -0.048942  -0.080237  0.015582  -0.056532  -0.056328  0.7430

#Checking minimum or maximum yields (100kg/2.47 acre)
print("Minimum yield of crops is  (100kg/2.47 acre):", df["Yield (Quintal/ Hectare) "].min())
print("Maximum yield of crops is  (100kg/2.47 acre):", df["Yield (Quintal/ Hectare) "].max())

Minimum yield of crops is  (100kg/2.47 acre): 1.32
Maximum yield of crops is  (100kg/2.47 acre): 1015.45

#Checking minimum or maximum cost production for c2 scheme (per 2.47 acre)
print("Minimum cost production for c2 scheme(per 2.47 acre):", df["Cost of Production (`/Quintal) C2"].min())
print("Maximum cost production for c2 scheme(per 2.47 acre):", df["Cost of Production (`/Quintal) C2"].max())

Minimum cost production for c2 scheme(per 2.47 acre): 85.79
Maximum cost production for c2 scheme(per 2.47 acre): 5777.48

#Rename the data
df.rename(columns={'Cost of Cultivation (`/Hectare) C2':'CC'}, inplace=True)
df.rename(columns={'Cost of Production (`/Quintal) C2':'CP'}, inplace=True)
df.rename(columns={'Yield (Quintal/ Hectare) ':'Y'}, inplace=True)
#show the dataframe
df.head()
```

Fig -5: Yield of crops min/max values

7.2 Data Visualization

Data Visualization is a significant ability in applied insights and AI. Measurements really does without a doubt zero in on quantitative portrayals and assessments of information. Information representation gives a significant set-up of apparatuses for acquiring a subjective comprehension. This can be useful while investigating and getting to know a dataset and can assist with distinguishing designs, degenerate information, anomalies, and considerably more. With a little space information, information perceptions can be utilized to communicate and exhibit key connections in plots and graphs that are more instinctive and partners than proportions of affiliation or importance. Information perception and exploratory information investigation are entire fields themselves and it will suggest a more profound jump into a few the books referenced toward the end.

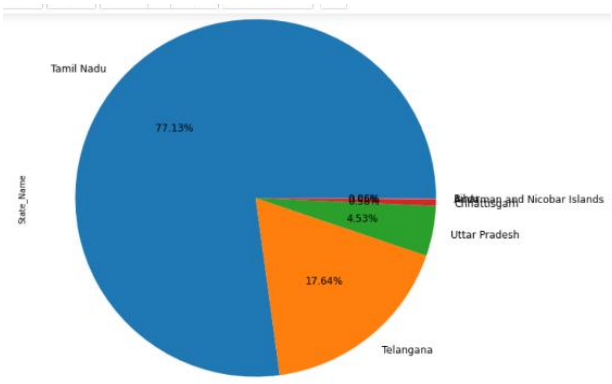


Fig -6: Crop production in various states

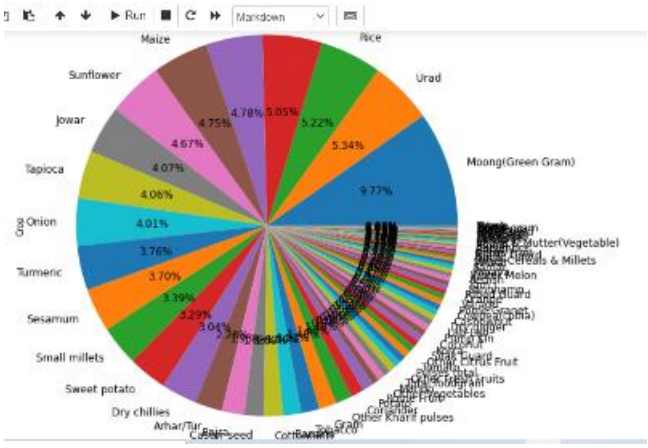


Fig -7: Proportion of Crops

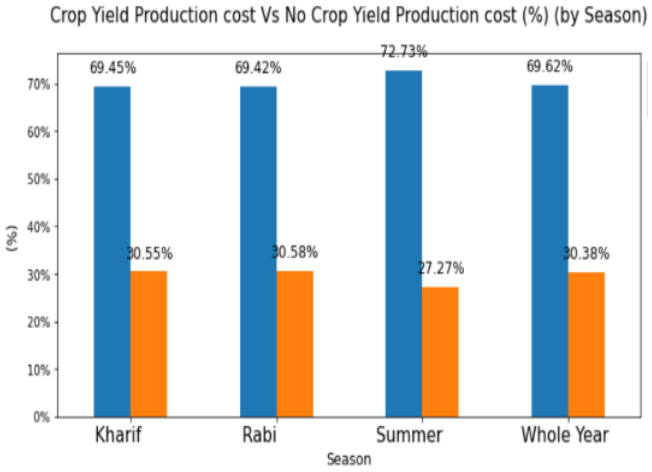


Fig -8: Crop Vs No Crop yield production cost

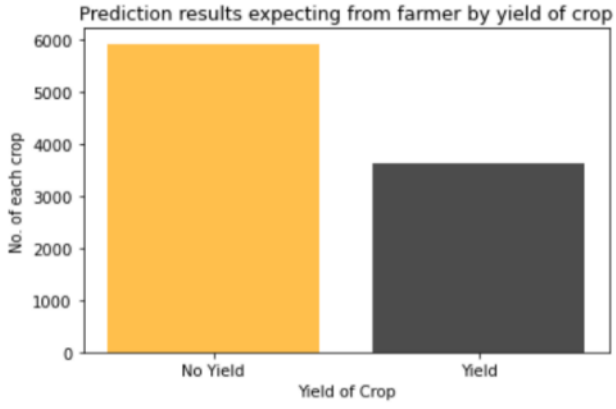


Fig -9: Prediction of yield of crop

7.3 Comparing Algorithm with prediction on basis of accuracy results

It is vital to analyze the exhibition of numerous different AI calculations reliably and it will find to make a test outfit to look at various changed AI calculations in Python with scikit-learn. It can involve this test bridle as a format on your own AI issues and add more and various calculations to analyze. Each model will have different execution qualities. Utilizing resampling techniques like cross approval, you can get a gauge for how exact each model might be on concealed information. It should have the option to utilize these appraisals to pick a couple of best models from the set-up of models that you have made. When have a new dataset, it is really smart to picture the information involving various strategies to check out at the information according to alternate points of view. A similar thought applies to show choice. You ought to utilize various perspectives on assessed precision of your AI calculations to pick the a couple to conclude. A method for doing this is to utilize different representation techniques to show the typical exactness, change and different properties of the dissemination of model correctness. In the example below 4 different algorithms are compared namely: Random Forest, Decision Tree Classifier, Naive Bayes and SVM. The K-overlay cross approval methodology is utilized to assess every calculation, critically designed with a similar arbitrary seed to guarantee that similar parts to the preparation information are performed and that every calculation is assessed in definitively the same way. Before that looking at calculation, Building a Machine Learning Model utilizing Scikit-Learn libraries. In this library bundle need to done preprocessing, straight model with calculated relapse strategy, cross approving by KFold technique, troupe with irregular timberland strategy and tree with choice tree classifier. Furthermore, parting the train set and test set. To anticipating the outcome by looking at precision.

7.3.1 Accuracy

The Proportion of the all out number of forecasts that is right in any case generally speaking how frequently the model predicts accurately defaulters and non-defaulters.

Accuracy can be calculated by:

(TP+TN) / (TP+FN+FP+TN), where TP, TN, FP, FN can be referred as True Positive, True Negative, False Positive, False Negative respectively. Accuracy is the most natural presentation measure and it is basically a proportion of accurately anticipated perception to the all perceptions. One might feel that, on the off chance that we have high exactness, our model is ideal. Indeed, precision is an incredible measure yet just when you have symmetric datasets where upsides of misleading positive and bogus negatives are practically same.

7.3.2 Precision

The proportion of positive predictions that are actually correct. **Precision = TP / (TP + FP),** precision is the proportion of accurately anticipated positive perceptions to the complete anticipated positive perceptions. The inquiry that this measurement answer is of all travelers that named as made due, what number of really made due? High accuracy connects with the low bogus positive rate. We have **0.788** accuracy which is very great.

7.3.3 Recall

The extent of positive noticed esteems accurately anticipated. (The extent of genuine defaulters that the model will accurately anticipate). **Recall = TP / (TP + FN).** Recall Sensitivity - It is the proportion of accurately anticipated positive perceptions to the all perceptions in genuine class - yes.

7.3.4 F1 Score

F1 Score is the weighted ordinary of Precision and Recall. Accordingly, this score considers both bogus up-sides and misleading negatives. Instinctively it isn't as straightforward as precision, yet F1 is normally more helpful than exactness, particularly assuming you have a lopsided class dissemination. Exactness works best

assuming bogus up-sides and misleading negatives have comparative expense. On the off chance that the expense of misleading up-sides and bogus negatives are altogether different, it's smarter to check out at both Precision and Recall.

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

F1Score=2*(Recall*Precision)/(Recall+Precision)

7.4 Deployment

7.4.1 GUI

Tkinter instructional exercise gives essential and high level ideas of Python Tkinter. Our Tkinter instructional exercise is intended for novices and professionals. Python gives the standard library Tkinter to making the graphical UI for work area based applications. Creating work area based applications with python Tkinter is anything but a complicated errand. A void Tkinter high level window can be made by utilizing the accompanying steps. Tkinter is a python library for creating GUI (Graphical User Interfaces). We utilize the tkinter interface and Tkinter will accompany Python as a standard bundle, it tends to be utilized for security reason for every clients or bookkeepers. There will be two sorts of pages like enlistment client reason and login section motivation behind library for mak-

ing a use of UI (User Interface), to make windows and any remaining graphical client clients.

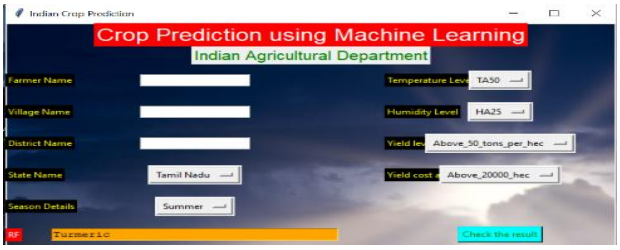


Fig -10: Crop Prediction

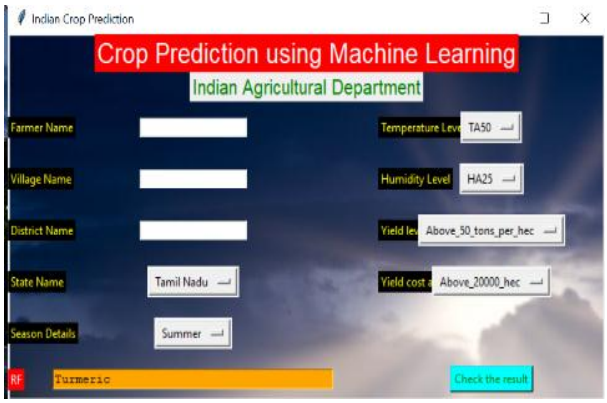


Fig -11: Cost Prediction

8. FUTURE ENHANCEMENTS

Remaining SMLT algorithms will be involved in finding the best accuracy with applying to predict the crop yield and cost. Agricultural department wants to automate the detecting the yield crops from eligibility process (real time). To automate this process by show the prediction result in web application or desktop application. To optimize the work to implement in Artificial Intelligence environment.

9.CONCLUSION

The insightful interaction began from information cleaning and handling, missing worth, exploratory investigation lastly model structure and assessment. At long last we foresee the yield utilizing AI calculation with various outcomes. This brings a portion of the accompanying experiences about crop forecast. As greatest kinds of yields will

be covered under this framework, rancher might get to be aware of the harvest which might in all likelihood never have been developed and rattles off every single imaginable harvest, it helps the rancher in decision making of which harvest to develop. Additionally, this framework thinks about the past creation of information which will assist the rancher with getting understanding into the interest and the expense of different harvests in market.

REFERENCE

[1] P. S. Maya Gopal and R. Bhargavi, "Feature selection for yield prediction in boruta algorithm," Int. J. Pure Appl. Math., vol. 118, no. 22, pp. 139–144, 2018.

[2] S. Ji, S. Pan, X. Li, E. Cambria, G. Long, and Z. Huang, "Suicidal ideation detection: A review of machine learning methods and applications," IEEE Trans. Comput. Social Syst., vol. 8, no. 1, pp. 214–226,Feb. 2021.

[3] K. Ranjini, A. Suruliandi, and S. P. Raja, "An ensemble of heterogeneous incremental classifiers for assisted reproductive technology outcome prediction," IEEE Trans. Comput. Social Syst.early access, Nov.3,2020doi: 10.1109/TCSS.2020.3032640.

[4] H. Liu and R. Setiono, "A probabilistic approach to feature selection-afilter solution," in Proc. 13th Int. Conf. Int. Conf. Mach. Learn., vol. 96,1996, pp. 319–327.