BEST SUITED CROP RECOMMENDATION SYSTEM USING ML

Presented By:

Bapatu Yeruguti Lakshmi Shreya - 190304105069
Konapalli Manjunadh   - 190304105084
Konuganti Manikanta Reddy   - 190304105085
Raneru Manoj Kumar   - 190304105097

Supervised by:

Suman Shekhar

OUTLINE

- Introduction

- Objectives of Project Work

- Literature Survey

- Problem Statement

- Implementation With Results (till date)

- Conclusion & Future Work

- Reference

INTRODUCTION

Agriculture plays an vital role in sustaining life, So setbacks in this could pose a serious danger. One such setback is not being able to produce right crop in certain given conditions.

By recommending the Best suited crops to be cultivated by farmers based on several parameters can help them to make an informed decision before cultivation. Not taking the right decision of what to cultivate is one of the possible causes for a lower crop yield for marginal farmers in India. The common problem existing among the Indian farmers are they don't choose the right crop based on their soil requirements. Due to this they face a serious setback in productivity. This problem of the farmers has been addressed through precision agriculture.

OBJECTIVES OF PROJECT WORK

- To develop a crop recommendation system, based on the dependent parameters to suggest the most suitable crop to be grown in certain conditions.

- To be able to shift towards Precision Agriculture from conventional Agriculture.

- To help farmers and agriculturists increase the crop yield and maximize the profits to be earned.

- The proposed system would help to minimize the difficulties faced by farmers in choosing a crop and maximize the yield.

LITERATURE REVIEW

Sl. No.	Paper Title	Year	Methods	Shortcomings
1.	Crop recommendation system for precision agriculture	2017	• The rules are generated in form of if-then rules • The rules generated from the above model is used to develop a recommendation system.	Less data used and collected dataset is limited to one district. Accuracy of models is less than ninety.
2.	Instance based random forest with rotated feature space	2013	• The rotated features are used to generate the Random Forest model and the decision trees are really sensitive to the rotation. • This rotation approach improves the diversity among the trees in the forest	Absence of study of the effect of different split mechanism of the feature space to the performance of the ensemble.

Sl. No.	Paper Title	Year	Methods	Shortcomings
3.	Ensemble based classification using small training sets : A novel approach	2014	• Generating a core set of instances or representative set that can be used to train a classifier with very small number of instances without loosing generality	The method is generic may enough to apply to data sets with high number of dimensions also
4.	A Novel Consistent Random Forest Framework: Bernoulli Random Forests	2018	• The training data set D is first partitioned into a Structure part and an Estimation part for achieving the consistency property of the proposed BRF. • The key factor lies in the two Bernoulli driven/controlled tree construction processes.	Training performance is significantly improved by the application of the filters. But, the test results have not increased correspondingly
5.	Why the Naive Bayes approximation is not as Naive as it appears	2015	• Analyzing under what circumstances the NBA and associated NBC can be expected to be suboptimal and develop general diagnostics with which a problem can be examined.	At the level of two features, only intra-schemata error cancelations are visible

Sl. No.	Paper Title	Year	Methods	Shortcomings
6.	Reduct based ensemble of learning classifier system for real-valued classification problems	2013	• Rough set attribute reduction presents systematic method to get set of attribute subsets that do not lose distinguishing information in the original data.	No specific mention of how many base classifiers are required.
7.	Handwritten digit recognition using hoeffding tree, decision tree and random forests — A comparative approach	2017	• After the building of every tree, the data is passed to the entire tree and closeness, also known as proximity, is computed for every single pair of case.	one can conclude that both Hoeffding tree and Random Forest would be effective only within time constraint
8.	Optimized very fast decision tree with balanced classification accuracy and compact tree size.	2011	• Optimized-VFDT algorithm that uses an adaptive tie mechanism to automatically search for an optimized amount of tree node splitting, balancing the accuracy and the tree size, during the tree-building process.	one can conclude that both Hoeffding tree and Random Forest would be effective but within time constraint

Sl. No.	Paper Title	Year	Methods	Shortcomings
9.	Crop prediction using predictive analytics	2021	• Predictive model often perform calculation during the transaction in order to evaluate the risk and opportunity of customer during the transaction.	NPK sensor used will detect only the nitrogen, potassium and phosphate present in the soil. Parameters like humidity, pressure, ph are not identified
10.	A study on various data mining techniques for crop yield prediction	2014	• To predict the preferences of the user from clickthrough data for personalized query obtained from users. • An algorithm used in architecture is much faster than the SpyNB algorithm	The dataset which is considered is small in some cases. The result can only be improved by using a large dataset.

RESEARCH GAP

- Less Emphasis on ensemble learning based approach.

- Use of Integrated data sets is minimal.

- Explanation behind the implementation of mathematical concepts used for various models is vague.

- Precision agriculture based on common parameters found in every geographical location is less used or mentioned.

PROBLEM STATEMENT

The change in environmental conditions due to modernized living standards has cause drastic setbacks especially in the field of agriculture. As a result, the crop that was suitable few years back in certain geographical location is not yielding the expected amount of production. The change in soil conditions and weather patterns are primary reason. This issue requires solution like Precision Agriculture where the most suitable crop could be predicted or recommended based on the dependent factors.
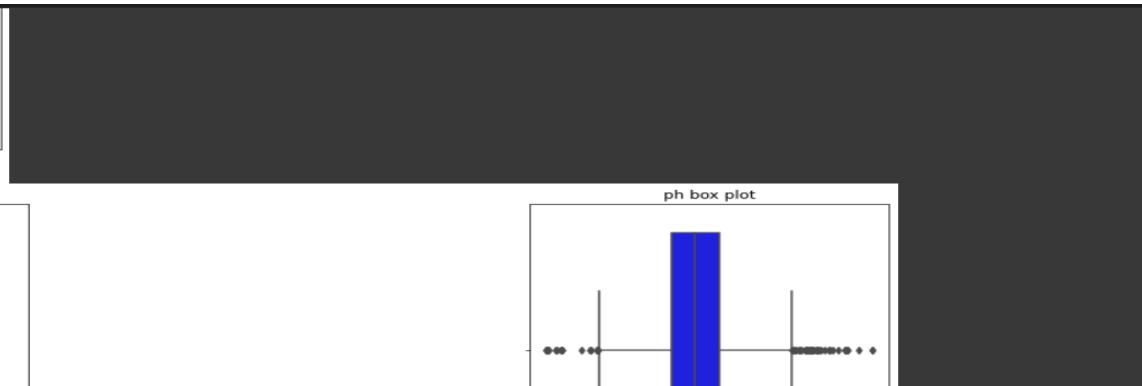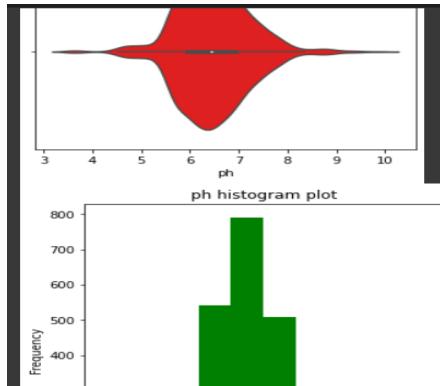
IMPLEMENTATION WITH RESULTS (TILL DATE)

- **Data Set** : Collected data set from source - Kaggle.
- **Data Cleaning** : Checked for redundant , null or any outliers found in the data and removed them using **NumPy** and **Pandas** libraries.

```
[6]   df.isnull().sum()
```

```
# Removing Outliers if any
Q1 = df.quantile(0.25)
Q3 = df.quantile(0.75)
IQR = Q3 - Q1
```

- **Data Visualization** : Used **Matplotlib** and **Seaborn** library so far to interpret and represent data using various graphs like histogram, box plot etc.

- **Machine Learning Algorithms** : Implemented Machine Learning Algorithms like Decision Trees, Naïve Bayes, SVM (support vector machine), Logistic Regression, Random Forest using **scikit-learn** library.

- **Ensemble Learning model** is used comprising of ml algorithms like Naïve Bayes, Random Forest, Decision Trees, Logistic Regression to improve the accuracy and efficiency of the system.

```
from sklearn.ensemble import VotingClassifier
vc=VotingClassifier(estimators=[ ('rf',RF),('gnb',Naive_Bayes), ('dt',DecisionTree) ,('lg', LogReg) ,('KNN', classifier) ,('svm',SVM)])
```

```
Ensemble Model-Voting Classifier Accuracy is 98.86363636363636
```

- **Wireframing** : Wireframing and layout design is done for the famers (client side portal)

TOOLS AND TECHNOLOGY

<u>Technology Used</u>- Machine Learning

<u>Programming Language Used</u>- Python

<u>Tools</u>:

- Anaconda (Jupyter Notebook), Google Collab

- Python Libraries: Pandas, Numpy, Scikit-learn, Matplotlib, Seaborn

- Flask

- Django or Node.js

- MongoDB or MySQL

COMPARISON OF EXISTING ALGORITHMS

Algorithm	Accuracy (%)	Description
SVM (Support Vector Machine)	10	The dimension of the hyperplane depends upon the number of features. It becomes difficult to imagine when the number of features exceeds three.
Random Forest	98.5	consists of a large number of individual decision trees that operate as an ensemble
Decision Tree	90.5	Decision trees can handle high-dimensional data.
Logistic Regression	95	Estimates the probability of an event occurring, based on a given dataset of independent variables.
Naive Bayes	90.5	Naïve Bayes is a probabilistic algorithm based on the Bayes Theorem, used in a wide variety of classification tasks.

CONCLUSION & FUTURE WORK

This project recommends the type of crop to be grown at certain place (Precision Agriculture) based on conditions such as whether, temperature, soil moisture and pH values etc.

• we are using Ensemble Learning model, but would be considering to tune the parameters of the model (Model Tuning) to check if the accuracy can be further improved.

• Along with it developing client side portal and integrating the ML model into it and hosting it over the internet.

REFERENCES

[1] H. Yang and S. Fong, *"Optimized very fast decision tree with balanced classification accuracy and compact tree size," The 3rd International Conference on Data Mining and Intelligent Information Technology Applications, 2011, pp. 57-64.*

[2] Y. Wang, S. -T. Xia, Q. Tang, J. Wu and X. Zhu, "A Novel Consistent Random Forest Framework: Bernoulli Random Forests," *in IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 8, pp. 3510-3523, Aug. 2018, doi: 10.1109/TNNLS.2017.2729778.*

[3] C. R. Stephens, H. F. Huerta and A. R. Linares, "Why the Naive Bayes approximation is not as Naive as it appears," *2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA), 2015, pp. 1-6, doi: 10.1109/IISA.2015.7388083.*

[4] S. Sahu, M. Chawla and N. Khare, "An efficient analysis of crop yield prediction using Hadoop framework based on random forest approach," *2017 International Conference on Computing, Communication and Automation (ICCCA), 2017, pp. 53-57, doi: 10.1109/CCAA.2017.8229770.*

[5] P. S. Vijayabaskar, R. Sreemathi and E. Keertanaa, "Crop prediction using predictive analytics," *2017 International Conference on Computation of Power, Energy Information and Commuincation (ICCPEIC), 2017, pp. 370-373, doi: 10.1109/ICCPEIC.2017.8290395.*

[6] B. Nassif, M. Azzeh, L. F. Capretz and D. Ho, "A comparison between decision trees and decision tree forest models for software development effort estimation," *2013 Third International Conference on Communications and Information Technology (ICCIT), 2013, pp. 220-224, doi: 10.1109/ICCITechnology.2013.6579553*

[7] R. Gupta et al., "WB-CPI: Weather Based Crop Prediction in India Using Big Data Analytics," *in IEEE Access, vol. 9, pp. 137869-137885, 2021, doi: 10.1109/ACCESS.2021.3117247.*

[8] N. Guntamukkala, R. Dara and G. Grewal, "A Machine-Learning Based Approach for Measuring the Completeness of Online Privacy Policies," *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), 2015, pp. 289-294, doi: 10.1109/ICMLA.2015.143.*

[9] R. Ye, Z. Le and P. N. Suganthan, "K-nearest neighbor based bagging SVM pruning," *2013 IEEE Symposium on Computational Intelligence and Ensemble Learning (CIEL), 2013, pp. 25-30, doi: 10.1109/CIEL.2013.6613136.*

[10] S. Zhang, "Challenges in KNN Classification," *in IEEE Transactions on Knowledge and Data Engineering, doi: 10.1109/TKDE.2021.3049250.*

[11] S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika and J. Nisha, "Crop recommendation system for precision agriculture," *2016 Eighth International Conference on Advanced Computing (ICoAC), 2017, pp. 32-36, doi: 10.1109/ICoAC.2017.7951740.*

[12] L. Zhang, Y. Ren and P. N. Suganthan, "Instance based random forest with rotated feature space," *2013 IEEE Symposium on Computational Intelligence and Ensemble Learning (CIEL), 2013, pp. 31-35, doi: 10.1109/CIEL.2013.6613137.*

[13] E. Debie, K. Shafi, C. Lokan and K. Merrick, "Reduct based ensemble of learning classifier system for real-valued classification problems," *2013 IEEE Symposium on Computational Intelligence and Ensemble Learning (CIEL), 2013, pp. 66-73, doi: 10.1109/CIEL.2013.6613142.*

[14] C. V. Krishna Veni and T. S. Rani, "Ensemble based classification using small training sets : A novel approach," *2014 IEEE Symposium on Computational Intelligence in Ensemble Learning (CIEL), 2014, pp. 1-8, doi: 10.1109/CIEL.2014.7015738.*

[15] K. Lavanya, S. Bajaj, P. Tank and S. Jain, "Handwritten digit recognition using hoeffding tree, decision tree and random forests — A comparative approach," *2017 International Conference on Computational Intelligence in Data Science(ICCIDS), 2017, pp. 1-6, doi: 10.1109/ICCIDS.2017.8272641.*

[16] K. Netti and Y. Radhika, "A novel method for minimizing loss of accuracy in Naive Bayes classifier," *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2015, pp. 1-4, doi: 10.1109/ICCIC.2015.7435801.*

[17] B. W. Yohanes, S. Y. Rusli and H. K. Wardana, "Location prediction model using Naïve Bayes algorithm in a half-open building," *2017 4th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), 2017, pp. 15-19, doi: 10.1109/ICITACEE.2017.8257667.*

[18] S. Zhang and J. Li, "KNN Classification with One-step Computation," *in IEEE Transactions on Knowledge and Data Engineering, doi: 10.1109/TKDE.2021.3119140.*

[19] K. Swati and A. J. Patankar, "Effective personalized mobile search using KNN," *2014 International Conference on Data Science & Engineering (ICDSE), 2014, pp. 157-160, doi: 10.1109/ICDSE.2014.6974629.*

[20] Y. Gandge and Sandhya, "A study on various data mining techniques for crop yield prediction," *2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), 2017, pp. 420-423, doi: 10.1109/ICEECCOT.2017.8284541.*

Thank you