

Efficient Plant Recommendation system using Random Forest model built on NodeJS stack

Jagadabi Chandrasekhar Varma ^[1], Koganti Aditya ^[2]

^{[1],[2]} CSE, VIT University - Vellore

ABSTRACT

In this project, we will be developing a Web Application to provide a user with Plant recommendations suiting their environmental conditions. We will primarily be making use of PyFlask, Random Forest Algorithm in our project, combined with a backend and a frontend developed using the AngularJS framework and MEAN Stack. We will use PyFlask to create the API to access the AI Model and then use a hosting service to host the same, then use the Random Forest Algorithm to be able to efficiently classify our data and provide us with relevant results related to the filters provided to the API. When a user visits the website and enters the filters related to their environmental conditions, then the data is sent to an API, on which the data is analyzed on the basis of these filters, and then the related items will be classified using Random Forest and the most accurate Plant Family Name will be returned and then all the plants' data from that Plant Family will be displayed to the user in a readable manner.

Keywords — AI, ML, Classifiers, Random Forest, Recommendation System

I. INTRODUCTION

Farming has been an important and major occupation in India and also in the whole world, it can play a part in determining the country's status and development and thus it is vital importance to look over and possible advance the agricultural capability as due to increasing demand and varying climates and environments. Not only that Even though many areas are converting from rural to town and then city, people also have a tendency seek relaxation, and nature is one of the main things there is missing in their city life, thus they lean towards raising a plant/plants and nurture over them. For all actions we are doing, ultimately what we seek is for the development of a plant to its best possible nature.

But everything is not our hands, we can take utmost care and put all our hard work into raising the plants, but we can't predict what we are doing will be useful or can we even worse the plant's health, there is a wide range of parameters we have to consider, such for example take the temperature, where some plants are best raised under cooler conditions, and fynbos plants it is an indigenous plant that grows exclusively in South Africa for its required conditions wind speed and humidity on a particular scale, there are others in this region which are extremely endemic forming 7th world's largest floral region. Thus for the factors such as these, we have to into consideration before farming or raising a plant and a huge proportion of farmers lack sufficient knowledge on these, and while the same is the case for a person who knows nothing but Jan ob life in an urban city.

Thus we determined to implement a web-based application in which people seeking information who seeks to know and

raise plants specific to their living environment a user in this portal, in which the user will be provided with

various filtering options, to determine the environmental conditions of their location, and the quality of lifestyle the plants can be provided with. The environmental factors include Climatic Conditions such as Temperature, the scope of the sun, etc., then Soil conditions, and the minimum area required for a plant to survive. Hence, based on these filters, an AI algorithm will be used. Along with using our application for Predicting the most suitable plants for our Environmental Conditions, the user can also purchase plants from the application itself.

A. THEORETICAL BACKGROUND:

Our project mainly relies on two parts, one is the UI where the user has the capability to interact, and this is achieved through the MEAN stack based on NodeJS, while the data that is requested/supplied by/to the user will be from the machine learning model which sends the data based the conditions the user requested, but it requires a data on huge quantity to train along with various parameters to consider. We will primarily be making use of PyFlask, Random Forest Algorithm in our project, combined with a backend and a frontend developed using the AngularJS framework and MEAN Stack. We will be PyFlask to create the API to access the AI Model and then use a hosting service to host the same, then use the Random Forest Algorithm to be able to efficiently classify our data and provide us with relevant results related to the filters provided to the API. When a user visits the website and enters the filters related to their environmental conditions, then the data is sent to an API, on which the data is

analyzed on the basis of these filters, and then the related items will be classified using Random Forest and the most accurate Plant Family Name will be returned and then all the plants' data from that Plant Family will be displayed to the user in a readable manner.

B. MOTIVATION:

There has been a Tremendous advancement in Technology these days, numerous technologies are coming into the light, correlated with real-world problems showing the solution, even still after all this, there should be significant use of such technology in the agricultural department, especially in India. People, specifically rural's, are very lacking in knowledge of tremendous information we were able to gather these years, they have been relying upon their ancestral knowledge and are hesitant to try something new upon their lands. Thus our idea behind the project came into existence.

C. AIM OF THE PROPOSED WORK

we aim to develop tools and services to serve a farmer's needs. One of which is to provide the user with Plant/crop recommendations suiting their environmental conditions. The environmental conditions submitted are analyzed using an ML model, and the most suitable species is to be displayed to the user. The requirements of casual gardeners and full-time farmers are not the same. Hence a Plant Recommendation will be developed alongside a Crop Recommendation system to separately cater to the needs of gardeners and farmers

D. . OBJECTIVE(S) OF THE PROPOSED WORK

Ongoing through the research papers, we understood that the Random Forest method is the one that suits our purpose the best. To say precisely, the filter method of the rf algorithm's attribute can be done using by using various similar procedures such tests, analyzing the variance, data extraction, etc. Random forests are known to be a variance of ensemble procedure that resorts to making predictions by averaging over the predictions of several independent base models. Since its introduction by Breiman in the year 2001, the framework of the random forest has been found to be extremely successful and useful as a general-purpose classification and regression method. So, we have developed an Application using MEAN Stack powered by AI consisting of the Random Forest Classifier.

II. OVERVIEW OF THE PROPOSED SYSTEM

E. Decision Tree:

A decision tree is an intuitive model and is the building block of a random forest. A decision tree can be thought of as the answers to our data to a series of questions with a yes/no which leads to a predicted class (in the case of regression, leads to a continuous value). Decision tree models can be called interpretable because of the nature of classification it performs which is similar to a human being's decisions: we

arrive at a decision after asking multiple questions and queries about the data that is available to us in an ideal world scenario.

F. Random Forest:

The Random Forest Algorithm is a Supervised Learning algorithm. Random Forest is similar to a decision tree, but as its name suggests 'Forest', it is nothing but a collection of multiple Decision Trees, which are combined to create a more precise and stable expectation when contrasted with a solitary Decision Tree. Then, this Random Forest can be used in Classification problems, such as this one.

G. Advantages of Random Forest Algorithm:

1. It can be used to solve multiple tasks such as the problems of Regression as well as Classification, which make up almost the maximum percentage of total ML.
2. It is an exceptionally Handy calculation since the default hyperparameters that it uses to deliver a decent forecast outcome.
3. It also partially solves the problem of overfitting in ML models.

H. Limitations of Random Forest Algorithm:

Random forest has that can become slow and ineffective for a large no. of trees and real-time predictions. The algorithms are generally easily trainable as the training process is fast but once they are trained they become very slow to create predictions. A larger number of trees in a decision tree gives more accurate results which leads to a lower model in terms of efficiency. These are fast in the most real-world applications but there were certain times in which run-time performances are more prioritized so alternative approaches are preferred in such cases.

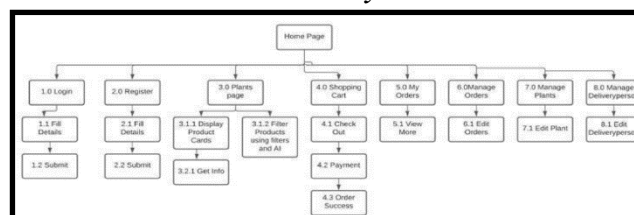
III. PROPOSED SYSTEM MODEL

A. Framework, Architecture, or Module for the Proposed System

H.A.1 MVC SUPPORT:

One of the best reasons for proposing node.js is that it supports the MVC pattern. This is useful for working on each area separately, such as, as the term implies models, views, and control, we can work independently without too much of each component influencing the other. The model module works as the blueprint of database structure and the view component represents the front-end part loaded based on the front-end frameworks from a database or templating engines while the controller part as the name implies it manages the control over all other components where it contains all the server-side logic such as maintaining the database and requests from a user from client-side and all business logic

H.A.2 Hierarchical Task Analysis:



H.A.3 State Transition Network for User:

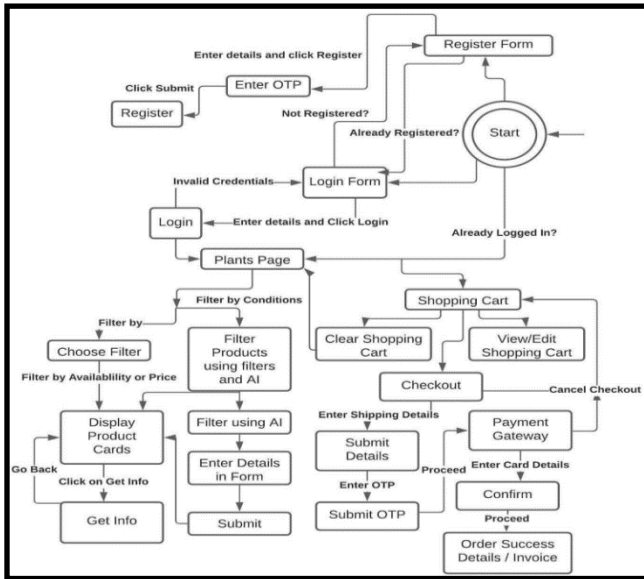


Figure 1

H.A.4 State Transition Network for Admin:

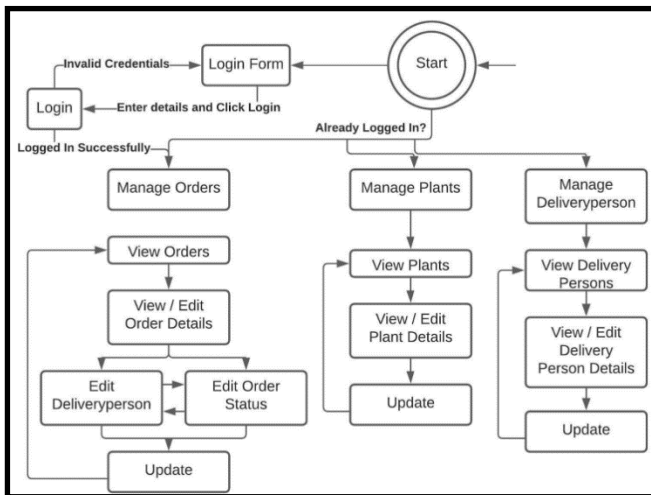


Figure 3

I. Architecture

we have chosen Random Forest as it starts from the top and divides the dataset into smaller decision trees. And then start combining from the bottom nodes, whereas in Boosting techniques they start dividing and combining from top to bottom itself. So, we have chosen Random Forest because it creates many trees each with leaves of equivalent load inside the model, hence providing better accuracy, whereas the boosting techniques introduce leaf weighting in order to penalize all the nodes that tend to reduce the model predictability

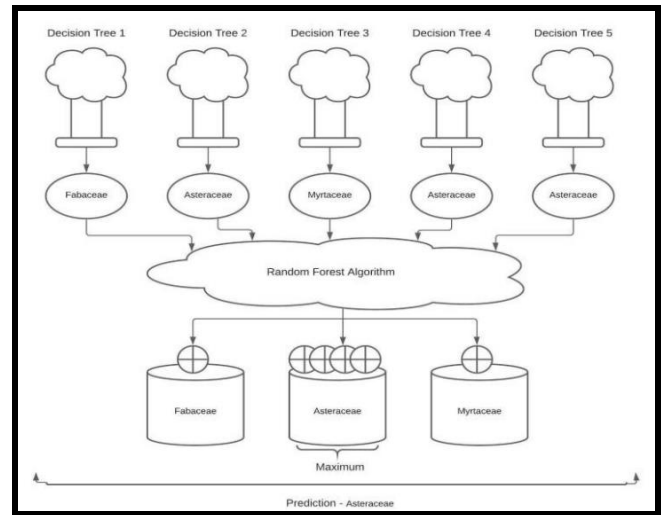


Figure 2: AI Model Implementation

J. Requirements:

J.A.1 Product Perspective

- As these deals with testing what it does rather than how it does, the required people will be farmers or people who raise plants for their hobby lacking knowledge
- also, people seeking the online trade.

J.A.2 Product Perspective

- Suggest plants/Crops based on the conditions given or required by the user
- Able to make online purchases
- capable of delivery through online purchases
- modern UI using NodeJs along with latest secure technologies to store data such as MongoDB for data and customer information

J.A.3 User Characteristics

- users who are seeking recommendations on plants they rise or crop they want to grow
- users who wish to make online purchases and delivery without any additional processes
- people seeking for employment as a delivery employee

J.A.4 Assumption & Dependencies

since our project's main functionality is recommendation/suggestion, and we will achieve it using machine learning and it requires two components

- Assumptions - if the data is not perfect no matter what other component is, we couldn't achieve perfect result so that would be one
- and the other is dependency where the model we used plays a significant impact, as without its suitable model for the project it also archives nothing regardless of how better the model is.

J.A.5 Domain Requirements

- no special requirements are needed to maintain the domain
- in case of online deployment, special charges may apply based on the portal we choose to, free are available such as Heroku, netlify, etc,
- Along with the above, the same is the case for mongo cloud for analysis or user analysis and product data analysis and also more additional benefits.

J.A.6 User Requirements

There are a total of 3 types of users this project is directed at

- Customer:** where a user who wishes gain certain knowledge on plants and buy, this user require:
 - login authentication
 - along with financial requirements for payment
- Delivery:** As same as a customer he needs authentication to prove his identity along with approval from the admin
- Admin:** finally, Admin, is only can be authorized by another admin, where he requires admin permission to access to change the source data for authenticating

IV. MODULES IN FRONT-END WEB APPLICATION

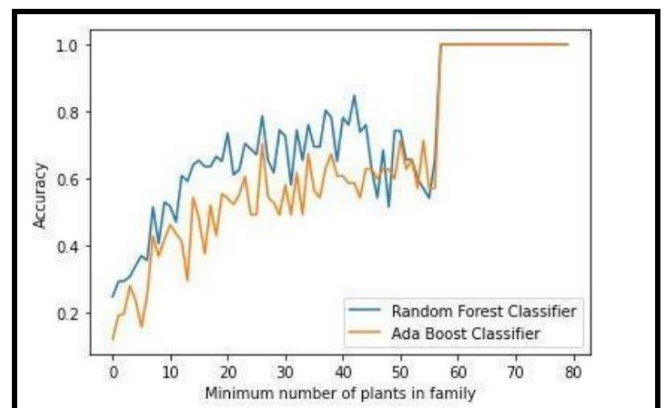
The various functional modules that are implemented in our project are:

- **Registration Module:** Every single customer needs to sign up once they visit "plantation.com".
- **Login Module:** Customers and Admins have to log in to the site with a valid email and password for shopping. Admin is given specific privileges once they log in.
- **Manage Plants, Delivery persons, and Orders:** Here, the admin has the privileges to add/edit or delete the Plants/Delivery persons/Orders from the display menu.
- **Filter Module:** Both the customer and admin will be able to filter/eliminate the items to narrow the items displayed.

- **Shopping Cart:** This means the customer after login is enabled to make orders or cancel orders of the products from the shopping cart. He/she may also change the item count of any products to proceed to checkout.
- **Payment Verification:** Before proceeding to the payment, a verification is done as a safety measure using OTP sent to the user's Email.
- **Payment Module:** In the system, we are dealing with the online mode of payment by card and Wallet. In this module, a Credit Card Validation algorithm by Luhn called Luhn's Formula has been implemented in order to avoid any Invalid Credit Cards to be used.
- **Log-out:** customers can log out after searching or ordering the products.
- **Bill Report Generation:** A copy of the bill with sent to the customer's email and another will send to the system database for auth purposes after ordering the products.
- **Conditions Specification Module (AI):** This is the module that allows the user to enter the Environmental Conditions for the plants they want, and then the user shall submit the form, then the contents run through the AI Model and the plant family related to the Conditions is returned and all the plants in the Plant Family are displayed.

IV. RESULTS AND CONCLUSION

Before removing families with less plants in them, we shall compare the accuracy score of the Random Forest, ADABOOST and SVM Classifier. In order to do this, we shall plot a graph for Random Forest and ADABOOST



Classifier by comparing the Minimum number of plants in a family vs Accuracy. i.e. finding the accuracy when all plant families with a count below 1 are removed, all the way up to 81.

To conclude, we can say that we have learned a lot during the course of this project and have successfully trained an AI

Model for predicting plants to the user which shall survive in their environmental conditions, and this has been tested as well. This project shall help all the users who want to grow plants in their houses or gardens, but do not know the suitable plants for their Environmental Conditions. Hence, they can now use this well-trained and tested Web Application to enter their Environmental Conditions, which then shall be analyzed by our AI Model, and which in return predicts the most suitable Plant Family for the input parameters.

FUTURE SCOPE

This application can be published online for the benefit of all users. And also in the application, we can add a feature which predicts plants based on the last purchased items, so that the user does not need to input the environmental parameters again and again.

REFERENCES

- [1] Jaiswal, J. K., & Samikannu, R. (2017, February). **Application of random forest algorithm on feature subset selection and classification and regression**. In 2017 World Congress on Computing and Communication Technologies (WCCCT) (pp. 65-68). IEEE.
- [2] Biau, G. (2012). **Analysis of a random forests model**. *The Journal of Machine Learning Research*, **13**(1), 1063-1095
- [3] Verma, A. (2018). **Study and Evaluation of Classification Algorithms in Data Mining**. *International Research Journal of Engineering and Technology (IRJET)* Volume, 5
- [4] Kumar, V., Dave, V., Bhadauriya, R., & Chaudhary, S. (2013, January). **Krishimantra: agricultural recommendation system**. In Proceedings of the 3rd ACM Symposium on Computing for Development (pp. 1-2).
- [5] Ajesh, A., Nair, J., & Jijin, P. S. (2016, September). **A random forest approach for rating-based recommender system**. In 2016 International conference on advances in computing, communications and informatics (ICACCI) (pp. 1293-1297). IEEE.
- [6] **Random forest classifier for multi-category classification of web pages** Publisher: IEEE Cite This PDF Win Thanda Aung; Yangon Myanmar; Khin Hay Mar Saw Hla (2009)
- [7] Ali, J., Khan, R., Ahmad, N., & Maqsood, I. (2012). **Random forests and decision trees**. *International Journal of Computer Science Issues (IJCSI)*, **9**(5), 272
- [8] Pretorius, A., Bierman, S., & Steel, S. J. (2016, November). **A meta-analysis of research in random forests for classification**. In 2016 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech) (pp. 1-6). IEEE
- [9] Paul, A., Mukherjee, D. P., Das, P., Gangopadhyay, A., Chintia, A. R., & Kundu, S. (2018). **Improved random forest for classification**. *IEEE Transactions on Image Processing*, **27**(8), 4012-4024.
- [10] Mustika, W. F., Murfi, H., & Widyaningsih, Y. (2019, October). **Analysis Accuracy of XGBoost Model for Multiclass Classification-A Case Study of Applicant Level Risk Prediction for Life Insurance**. In 2019 5th International Conference on Science in Information Technology (ICSITech) (pp. 71-77). IEEE..
- [11] An, T. K., & Kim, M. H. (2010, October). **A new diverse AdaBoost classifier**. In 2010 International Conference on Artificial Intelligence and Computational Intelligence (Vol. 1, pp. 359-363). IEEE.
- [12] Chengsheng, T., Huacheng, L., & Bing, X. (2017). **AdaBoost typical Algorithm and its application research**. In MATEC Web of Conferences (Vol. 139, p. 00222). EDP Sciences