Plant Recommendation Model for Urban People Using Majority Voting Classifier

ANANTHA LAKSHMI M[1] DIVYA PRIYANKA V[2] NOOKAMBIKA S[3]
IMRAN FAZIL Sk[4] SUMANTH V[5]

[1,2,3,4,5] Department of Computer Science and Engineering

Sasi Institute of Technology and Engineering, Tadepalligudem, Andhra Pradesh, India, 534101

anantha@sasi.ac.in divya5g7@sasi.ac.in nookambika5d4@sasi.ac.in

imran5e0@sasi.ac.in sumanth5h0@sasi.ac.in

Abstract— **There is no scarcity of fresh air in villages and gardens but there is a vast necessity for pure air in cities. People from villages majorly do know well about plants but the majority of the people who stay in cities and who have intentions to grow plants won't have enough prior knowledge of the selection and classification of plants based on growing parameters. The main challenge would be choosing the correct plant for the correct season. Having a habit of growing or maintaining a garden is good but the real challenge is how long the zeal of growing the plant species rests in mind if the situations are odd or against the suitable habitation? This proposed method aims to build a recommendation system for plants based on parameters for urban people using majority voting classifier which helps the gardeners to gain more interest in purifying the air We believe that the plants are natural air filters, if there are more plants then the amount of fresh air is produced is also more. The parameters considered in this model are the name of the plant, season, minimum and maximum temperature, minimum and maximum humidity, and the soil type is the growing conditions for plants. The dataset considered in this model is our customized dataset which consists of fruits, flowers, vegetables, and herbs. This recommendation model recommends the plants based on the user's input.**

Keywords— Machine Learning, Recommendation Systems, Decision Tree, Plant Recommendation, Majority voting classifier

I. INTRODUCTION

Plants are essential to almost all other forms of life since they make up the majority of living beings capable of converting sunlight into nourishment. Furthermore, because plants produce nearly all of the oxygen in the air that people and other animals breathe, it is difficult to imagine human life on Earth without plants. Plant classification aids in the preservation and survival of natural life. The ability to recognize plant species also helps to find out the plant population and its distribution. Plant recognition or classification can be done based on parts of the plant like leaves, flowers, and fruits of the plant. Many types of research are based on leaves in plant recognition, because the leaf was the most important part of the plant carrying its characteristics compared with other plant parts, fruits and flowers are not available throughout the year, most plants are seasonal as well as shape, size, and color of fruits and flowers are changing during growth. Many studies used leaves to identify plant categories based on shape, texture information, venation, and color. Most plants have distinctive leaves that differ from one another based on a variety of factors such as shape, color, texture, and margin. Because the substantial information contained by each can be utilized to identify and categorize the plant's origin or kind, leaf recognition/classification is a critical step in the plant classification process. This chapter briefly describes the introduction part of this work.

Seasonal plants are those that grow and flower during a specific season. If this plant is grown in an improper season, it will become undersized and easily affected by the disease. Some flowering plants bloom throughout the year and are unaffected by full shadow or full sun conditions nevertheless, this type of plant requires special care. The sunflower is an annual flowering plant that blooms throughout the year.

The following is how this document is structured: Section II presents related works on recommendation system. Section III presents the proposed methodology. Section IV gives the outcomes of our experiments. Finally, Section V presents conclusions and further enhancement.

II. LITERATURE SURVEY

Various Plant methods were proposed and reviewed here. The following section summarizes the history of those works which are done previously, highlighting the strengths and weaknesses of each method.

- Plant taxonomy
- Validation of plant species
- Combined plant classification
- Automated plant identification
- Application based plant identification
- Multi-organ plant identification

A. Plant taxonomy

Luciano D.S. Pacifico et al. (2018) [4] implemented a model using Multi-Layer-Perceptron (MLP) artificial neural network trained with Back propagation algorithm to perform automatic plant classification. Comparison of five different classifiers from plant classification literature and some of their variants is performed. They are: Decision Tree classifier (DT), Naive Bayes classifier (NB), K- Nearest Neighbors (KNN, with k = 3, 4 and 5), Support Vector Machine with RBF (SVM rbf) and Linear (S V M linear) kernel functions and a Multi-Layer Perceptron trained with Back propagation algorithm (MLP-BP). All these algorithms are implemented in Python programming language and used 1000 epochs for MLP-BP training. For comparison purposes, three real-world plant data sets obtained from UCI Machine Learning repository are employed: Iris, Wheat Seeds and 100 Plant Leaves. 100 Plant Seeds data set have been divided in seven data sets, so the test could be made on each individual plant leaf feature and all possible combinations of the three features (leaf margin, leaf shape and leaf texture). The experiment's results showed that MLP-BP outperforms all other algorithms in terms of overall accuracy. The advantage of this methodology is it gives more accurate and reliable results by also including the decomposed leaves in the dataset. The drawback of this methodology is that it did not include the other plant features extracted automatically using image processing techniques, and smaller dataset is considered.

Xiao-Feng Wang et al. (2008) [25], came up with an efficient classification framework to classify leaf images with complicated background where some interferents and overlapping phenomenon may exist. To segment leaf images with complex backgrounds based on past shape knowledge, an automatic marker-controlled watershed method is used in combination with pee- segmentation and morphological operations. After watershed segmentation and leafstalk removal, twenty-three moment invariant are retrieved from binary images, comprising seven Hu geometric moments and sixteen Zernike moment features. In addition, to handle retrieved high-dimensional features, an efficient Moving Centre Hypersphere (MCH) classifier with data compression function is presented.

B. Validation of plant species

E M Imah et al. (2018) [8], implemented an automatic plant recognition based on digital leaf images. Plant species have distinct leaf traits such as form, texture, border, and colour that distinguish them from one another. GRLVQ is a competitive learning algorithm that combines the extraction and classification of features. The results of the experiments reveal that GRLVQ outperforms the preceding algorithm. The outcome indicates that GRLVQ outperforms the competition. GRLVQ has a precision and recall of 0.9380 and 0.9298, respectively, and performs well across the board. The precision and recall of the random forest model are 0.885 and 0.882, respectively, whereas the precision and recall of the SVM model are 0.911 and 0.634, respectively. This study uses full feature attributes to classify with Random Forest, SVM, and GRLVQ classification algorithms. GRLVQ outperforms random Forest and SVM in recognizing the five classes in the EEG epileptic seizure dataset, with an accuracy of 92.98 %, over 5% higher than random Forest and more than 25% better than SVM. The advantage of this methodology is that it has the quickest CPU time for plant species recognition but mathematical knowledge is needed to implement this system.

Amala Sabu et al. (2017) [11], proposed and implemented a system for automatic identification of medicinal plants from their leaves. This proposed system makes use of computer vision and machine learning to identify a pre-trained medicinal plant from its leaf. A combination of SURF and HOG features retrieved from leaf images, as well as a classification applying a k-NN classifier, are used. SURF is a type of image that is invariant to affine transformations such as scaling and rotation. The concept of integral image is used to extract the SURF feature. The gradient orientation in a small region of an image is counted using this technique. The usage of typical shape and colour traits of leaves, which are computationally difficult to extract because they are spatial features, is the work's key focus.

C. Combined plant classification

M Parul Mittal et al. (2018) [6], discussed the improved performance of the combined classifier technique for the classification and identification of plant leaf images. The research technique shows how to extract features such as shape, colour, texture, and vein. There were a total of 139 characteristics retrieved, which were then classified and identified using various classifiers such as SVM, Decision tree, and combination classifier. The results reveal that the accuracy varies depending on the amount of input leaf photos and classes used in the classification process, with the maximum accuracy being attained using the combination classifier. In the aforementioned results, the performance of the combined classifier was compared to that of the SVM, Nave Bay's, and Decision tree, demonstrating that the performance accuracy is increased when compared to existing techniques. This system is efficient and takes less computational time. The limitation of this model, is it identifies images only from the white background. It has to be improved to identify the category of leaf image from coloured background as well as in multiple object images.

Janez Demsar (2006) [28], discusses the reviews on the current practice and then theoretically and empirically examines several suitable tests. Based on that, a set of simple, yet safe and robust non- parametric tests for statistical comparisons of classifiers: the Wilcoxon signed ranks test for comparison of two classifiers and the Friedman test with the corresponding post-hoc tests for comparison of more classifiers over multiple data sets were recommended. Results of the latter were neatly presented with the newly introduced CD (critical difference) diagrams. The results shown should be carefully concluded otherwise it may lead to incorrect values. But statistical testing can be considered which may possibly even favoured over pure improvements in the predictive power.

D. Automated plant identification

Jana Waldchen et al. (2018) [3], Proposed Current rates of species extinction have prompted a slew of efforts to protect and conserve biodiversity. Species conservation, on the other hand, necessitates proficiency in species identification, which can only be attained via extensive training and experience. Accessible, up-to-date technologies automating the process of species identification would be extremely beneficial to field researchers, land managers, educators, civil employees, and the general public. Since the user is only required to take an image and browse through the best-matching species, the identification involves no effort on their part. Furthermore, only rudimentary expert knowledge is required, which is critical considering the continued scarcity of qualified botanists. Non-experts with minimal botanical training and skill can also contribute to the census of the world's biodiversity thanks to an accurate automated identification method. Hence, approaching trends and technologies like augmented reality, data glasses, and 3D scans provide a long-term study and application viewpoint for such applications. The disadvantage of this model is it takes incorrect data, so the accuracy is very less.
Taisong Jin1 et al. (2015) [14], Proposed the method using sparse representation of leaf teeth characteristics, a novel automatic plant species identification. For plant species identification, our proposed method added four leaf teeth traits. The plant species for a test sample was then identified using a sparse representation-based classifier. The studies were carried out on a real-world dataset, demonstrating that our proposed method could be used to identify plants. Also tells about the intend to investigate the more complicated aspects of leaf teeth in future research. The advantage of this model is that characteristics such as shape, colour, and texture will also be included in the sparse representation-based plant species classification.
Neeraj Kumar et al. (2012) [22], developed the world's first computer vision system for automatically identifying plant species. Several key aspects of this system rely on computer vision, including classifying images as leaves or not, obtaining fine-scale segmentation of leaves from their backgrounds, efficiently extracting histograms of curvatures along the contour of the leaf at multiple scales, and retrieving the most similar species matches using a nearest neighbour search on a large dataset of 184 trees of labelled images using a nearest neighbour search on a large dataset of 184 trees of labelled images.

E. Application based plant identification

Cixiao Wang. (2017) [9], this study used software to conduct research and applied the app to outdoor mobile learning. It used an education experiment and a classroom observation method. The generative rule system and the production rule system are two expert systems that are employed. The plant facts database of non-attribute rules in the generative rule system has an attribute value of "yes" or "no". Learners can use the production rule system to identify plants based on a range of plant traits. The plant fact information, which combines the functioning of the corresponding machine, is known as the interpreter. Learners can use the plant identification and learning app to answer system questions while also observing plant traits, which can considerably increase their plant knowledge learning and mastery. As a result, according to the findings, the app can improve students' attitudes toward natural science and boost their interest in learning related disciplines like plants. In the future, learners' communicative function can be improved by combining peer collaborative inquiry learning with outdoor experiential learning under the supervision of teachers. The disadvantage is built an app is more complex.

Zhong-Qiu Zhao et al. (2014) [16], to automatically identify plant species quite valuable for Ecologists, amateur botanists, educators, and others. The Leafsnap is the first effective mobile app that addresses this issue. The Leafsnap, on the other hand, is built on the IOS platform. This system performs well with cutting-edge identifying technology. Eclipse is the software development platform, and Java is the programming language. The disadvantage is that it is difficult to understand the process of this model.

F. Multi organ plant identification

Yu Sun et al. (2017) [10], proposed the first BJFU100 dataset, which contains 10,000 photos of 100 plant species and serves as a data pillar for future plant identification research. We're continuing to add to the BJFU100 dataset by including more species and seasons. The dataset is available to everyone in the academic community. The BJFU100 database will be expanded in the future to include more plant species at various stages of their life cycles, as well as more thorough annotations. From classification to yield prediction, insect detection, and disease segmentation, the deep learning model will be extended. The result of this model is 91.78% accuracy. The advantage of this model is the dataset is available to everyone in the academic community. Disadvantage of this model is it took a small data set.

Nuril Aslina Che Hussin et al, (2013) [20] recommended the five steps: Picture capture, image pre-processing, feature extraction, identification, and performance assessments. The phase of feature extraction is the most critical and crucial in plant identification. To identify plants, this work provides the Scale Invariant Feature Transform (SIFT) method of shape feature extraction and the Grid Based Colour Moment (GBCM) approach of colour feature extraction. The dataset contains 11572 images of 126 tree species from the French Mediterranean region. The accuracy of this proposed system was 87.5 percent. The advantage of this model is to boost the identification rate, more robust features such as texture can be added. We also plan to experiment with learning classifiers such as supervised classifiers. The disadvantage of this model is less accuracy.

III. METHODOLOGY

The methodology provides information about the system design. Here, the system architecture diagram of the proposed system has been discussed. It deals with the design of the proposed system. It includes the system architecture, to represent the design of the system.

This project has only a system architecture diagram. The implementation details like modules or steps in implementation and proposed approach, techniques, and equation. The proposed work, which is implemented in Python 3.6.4 with libraries NumPy, pandas, matplotlib, TensorFlow, sci-kit-learn, Keras, Label encoder, Train test split and other mandatory libraries. the standard modified and simplified Decision Tree algorithm was proposed. The data is taken our own customized data. The dataset consists of both training and testing data.

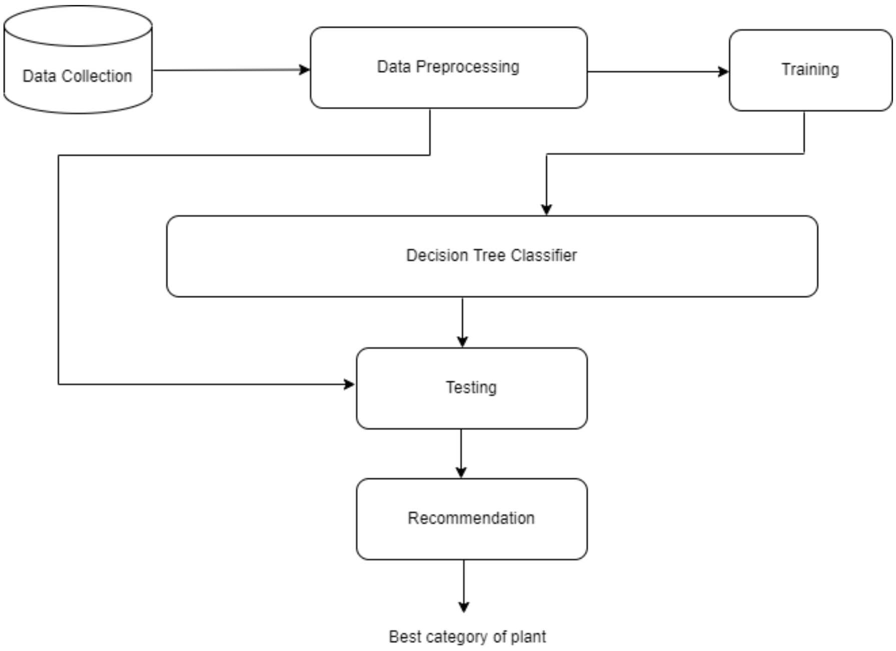

Figure 1. Proposed Architecture

This methodology gives information about the system design. Here, the system architecture diagram of the proposed system has been discussed. This chapter deals with the design of the proposed system. It includes the system architecture, to represent the design of the system. This project has only a system architecture diagram. The implementation details like modules or steps in implementation and proposed approach, techniques, and equations have been discussed.

A system architecture can be made up of a system as well as the sub-systems that will be built to implement the overall system. The above diagram represents the system architecture for recommendation and classification of plants using decision tree algorithm. In proposed methodology it contains two phases: training phase and recommendation phase. In the training phase, system is trained with the dataset which contains temperature, humidity, season and soil type details of the plants. Another phase is the recommendation phase, in which recommendation made based on the user's input. The recommendation is made based on the user's input. If the input is incorrect or inappropriate data then it shows error and exits the system. The dataset considered is used for training and testing. In this model, based on the user's input recommendation is divided into three types.

If the season is given as input then the model shows all the plants that can grow in that particular season as output which is independent of the area or the location of the user. By using this user gets to know which type of plants grows in that season. The second recommendation is if the inputs given are temperature and humidity then the plant that can be grown in that suitable conditions and the type of soil in which that plant grows are given as output.
The third type of recommendation is if the label or the plant name is given as input by user then the suitable conditions like season, temperature, humidity and soil type that are convenient for that plant to grow are given as output. Classification can also be made if all the conditions like plant name, season, temperature, humidity and soil type are given as input then it classifies as yes/no. If the model shows result as yes then the plant is suitable to grow otherwise no. Only the first type of recommendation is independent of the location remaining all other recommendations are dependent on the location given by user.

The proposed work is implemented in Python 3.6.4 with libraries scikit-learn, pandas, matplotlib, numpy, label encoder and other mandatory libraries. For this project a customised dataset containing some important parameters like temperature, humidity, season and soil are considered. These parameters are pre-processed and then machine learning algorithms are applied. Finally, the results are evaluated.

A. Modules of proposed system

The dataset collecting, dataset pre-processing, dataset splitting, and machine learning techniques are the four key modules of the proposed work. The live dataset will be gathered first from Twitter. The data will then be pre-processed, and the data will be divided into two datasets: train and test sets. Lastly, the results from the various machine learning algorithms are evaluated. The modules are given below:

- Dataset collection
- Data Pre-processing
- Splitting the data set
- Applying machine learning algorithm

1) Dataset Collection
The dataset considered in this model is our own customized dataset of 1340 species. It contains fruits, flowers, vegetables and herbs.

1	Name	Season	MinTemp	MaxTemp	Minhumid	Maxhumid	Soil
2	Mango	summer	23	30	50	95	black
3	Apple	winter	21	24	90	95	brown
4	Orange	winter	12	37	45	50	black
5	Promogranate	monsoon	32	41	80	85	light grey
6	Guava	spring	20	25	75	85	black
7	Grapes	spring	25	32	50	90	black
8	Banana	monsoon	13	38	65	80	brown
9	Custard Apple	winter	15	20	70	80	black
10	Watermelon	winter	21	32	60	80	dark brown
11	BlueBerry	winter	-25	-20	70	90	red
12	BlackBerry	winter	0	10	90	98	dark brown
13	Cherry	winter	30	45	26	30	drak brown
14	Coconut	rainy	27	72	80	85	dark brown
15	Papaya	monson	21	32	30	40	dark brown
16	Pineapple	summer	15	30	30	50	reddish brown
17	Lychee	summer	30	32	85	90	brown
18	Strawberry	monsoon	16	27	70	80	black
19	Sapota	winter	10	38	65	75	brown
20	Lemon	summer	25	40	50	60	black
21	Cashew Apple	winter	20	30	85	90	brown

Figure 2. Dataset (First 21 values are shown)

The parameters considered in the dataset are name of the plant, season, minimum and maximum temperature, minimum and maximum humidity, and soil type. These records are collected from the Google. Each and every plant is searched in Google with their minimum and maximum temperature, minimum and maximum humidity, and soil type.

2) Dataset Pre-processing

Data pre-processing is the process of transforming raw data into an understandable format. One can't work with raw data, thus this is a key stage in data mining. Before using machine learning or data mining methods, the quality of the data should be checked. The purpose of data pre-processing is to ensure that the data is of good quality. The quality of the data can be checked by following factors. The pre-processing task was implemented by a Pre-processor helper class that removes symbols and non-printing characters from documents, tokenizing it into words and removing meaningless stop-words. Here the purpose of pre-processing is to convert raw data into a form that fits to the machine learning techniques. The technique includes labelling, cleaning and data formatting. The dataset which is converted to CSV format is considered for further processing.

TABLE I. ENCODED VALUES FOR SOIL

Encoded Values for Soil	**Colour of Soil**
0	Black
1	Blue
2	Brown
3	Dark black
4	Dark brown
5	Light black
6	Light brown
7	Light grey
8	Light red
9	Light yellow
10	Loamy
11	Red
12	Reddish
13	Reddish Brown
14	Sandy
15	Yellow

TABLE II. ENCODED VALUES FOR PLANT CATEROGY

K value	**Plant category**
0	Flower
1	Fruit
2	Herb
3	Vegetable

TABLE III. ENCODED VALUES FOR SEASON

Encoded values for season	**Name of the season**
0	Autumn
1	Common
2	Fall
3	Monsoon
4	Rainy
5	Spring
6	Summer
7	Winter

3) Splitting the dataset

The train-test split is a technique for evaluating the performance of a machine learning algorithm. It can be used for classification or regression problems and can be used for any supervised learning algorithm. The procedure involves taking a dataset and dividing it into two subsets. The first subset is used to fit the model and is referred to as the training dataset. The second subset is not used to train the model; instead, the input element of the dataset is provided to the model, then predictions are made and compared to the expected values. This second dataset is referred to as the test dataset. Train Dataset: Used to fit the machine learning model Test Dataset: Used to evaluate the fit machine learning model.

The objective is to estimate the performance of the machine learning model on new data which is not used to train the model. The procedure has one main configuration parameter, which is the size of the train and test sets. This is most commonly expressed as a percentage between 0 and 1 for either the train or test datasets. For example, a training set with the size of 0.67 (67 percent) means that the remainder percentage 0.33 (33 percent) is assigned to the test set. There is no optimal split percentage. You must choose a split percentage that meets your project's objectives with considerations that include:

- Computational cost in training the model
- Computational cost in evaluating the model
- Training set representativeness.
- Test set representativeness

Common split percentages include:

- Train: 80%, Test: 20%
- Train: 67%, Test: 33%
- Train: 50%, Test: 50%

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state = 0)
In this proposed system, Train: 80% Test: 20% split percentage is used for applying machine learning algorithm such as Decision Tree machine is considered for learning and recommendation.

4) Applying machine learning algorithms
Machine learning is extremely complicated, and its operation differs based on the goal and the algorithm employed to complete it. A machine learning model, on the other hand, is a computer that looks at data and identifies patterns, then uses those insights to better execute its assigned task. Machine learning can automate any operation that relies on a set of data points or rules, including more complicated tasks like answering customer service calls and analysing resumes. Machine learning algorithms use more or less human intervention/reinforcement depending on the situation. supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning are the four major machine learning models.

4.1 Decision Tree
The most powerful and widely used tool for categorization and prediction is the decision tree. A decision tree is a flowchart-like tree structure in which each internal node represents an attribute test, each branch reflects the test's conclusion, and each leaf node (terminal node) stores a class label. A random forest contains many decision trees which gives much better result.

4.2 Constructing Decision Tree
By separating the source set into subgroups based on an attribute value test, a tree can be "trained." Recursive partitioning is the process of repeating this method on each derived subset. When all of the subsets at a node have the same value of the target variable, or when splitting no longer adds value to the predictions, the recursion is complete. Because the building of a decision tree classifier does not necessitate domain expertise or parameter selection, it is well suited to exploratory knowledge discovery. High-dimensional data can be handled via decision trees. The accuracy of the decision tree classifier is generally good. A common inductive strategy to learning classification information is decision tree induction.

B. Parameters

The Parameters used in this project are:

- Temperature
- Humidity
- Season
- Soil Type

1) Temperature
Temperature is a numerical value that indicates how hot or cold something is. When one body comes into contact with another that is colder or hotter, thermal energy is manifested, which is present in all matter and is the source of heat, a flow of energy. Temperature and heat should not be confused. The temperature is measured using a thermometer. Temperature scales that have traditionally defined temperature using a variety of reference points and thermometric substances are used to calibrate thermometers. The most common scales are the Celsius scale (previously known as centigrade, denoted as °C), the Fahrenheit scale (denoted as °F), and the Kelvin scale (denoted as K). The Kelvin scale (abbreviated K) is primarily used for scientific purposes under the International System of Units conventions (SI). Absolute zero is the lowest theoretical temperature at which no additional thermal energy from a body can be collected. The third law of thermodynamics acknowledges that it can only be approached with extreme precision (100 pK), but not reached experimentally. Temperature is important in many fields, including physics, chemistry, Earth science, astronomy, medicine, biology, ecology, material science, metallurgy, mechanical engineering, and geography, as well as most aspects of daily life.

Temperature is divided into minimum and maximum temperatures in this work. Temperature scales that have traditionally defined temperature using a variety of reference points and thermometric substances are used to calibrate thermometers.

2) Humidity
The amount of water vapour in the air is referred to as humidity. Water vapour, the gaseous state of water, is frequently invisible to the naked eye. [1] Humidity indicates the presence of precipitation, dew, or fog. The temperature and pressure of the system determine humidity. When the same amount of water vapour is present, cool air has a higher relative humidity than warm air. A related metric is the dew point. As the temperature rises, so does the amount of water vapour required to reach saturation. When a parcel of air's temperature falls below a certain threshold, it will eventually approach saturation without adding or losing water

mass. Water vapour concentrations in a given volume of air can vary dramatically. A parcel of near-saturated air may contain 28 g (0.99 oz) of water per cubic metre of air at 30 °C (86 °F), but only 8 g (0.28 oz) at 8 °C (46 °F).
The most common humidity measurements are absolute, relative, and specific. Absolute humidity is calculated using the mass of water vapour per volume of wet or the mass of water vapour per mass of dry air. The amount of water vapour in a given volume of air can vary dramatically. At 30 °C (86 °F), a parcel of near-saturated air may contain 28 g (0.99 oz) of water per cubic metre of air, but only 8 g (0.28 oz) of water per cubic metre of air at 8 °C (46 °F).

Absolute, relative, and specific humidity measurements are the most commonly used. The mass of water vapour per volume of wet air or the mass of water vapour per mass of dry air are used to calculate absolute humidity (usually in. [3] The term "relative humidity" refers to the difference between the current level of absolute humidity and the maximal humidity at the same temperature. The ratio of water vapour mass to total moist air parcel mass is known as specific humidity.

Humidity has a significant impact on surface life. High humidity reduces the rate of moisture evaporation from skin surfaces, which reduces heat exchange efficiency in animals that rely on perspiration (sweating) to control internal body temperature.
The idea of air "holding" or being "saturated" by water vapour is frequently emphasised in relation to the concept of relative humidity. This is mis-leading, because the amount of water vapour that enters (or can enter) a given region at a particular temperature is virtually completely independent of the amount of air (nitrogen, oxygen, and so on) present. Indeed, the equilibrium capacity to hold water vapour in a vacuum is roughly the same as the same volume filled with air; both are determined by the equilibrium vapour pressure of water at the particular temperature.

3) Season

A season is a division of the year that is determined by weather, ecology, and the number of daylight hours in a specific location. The Earth's orbit around the Sun and its axial tilt in relation to the ecliptic plane cause seasons. Seasons in temperate and polar regions are defined by variations in the amount of sunlight reaching the Earth's surface, which can cause animals to hibernate or migrate and vegetation to go dormant. Because different civilizations define the number and type of seasons based on regional differences, the number of seasons varies between present and historical cultures.

The Northern Hemisphere receives more direct sunlight as the hemisphere faces the Sun in May, June, and July. In the Southern Hemisphere, the same is true in November, December, and January. Because of the Earth's axial tilt, the Sun rises higher in the sky during the summer months, increasing solar flux. Due to seasonal lag, the Northern Hemisphere's warmest months are June, July, and August, while the Southern Hemisphere's warmest months are December, January, and February.

The four seasons recognised by the Gregorian calendar in temperate and sub-polar countries are spring, summer, autumn (or fall), and winter. Ecologists frequently use a six-season model for temperate climate regions with no fixed calendar dates: preverna1, vernal, estival, serotinal, autumnal, and hibernal. Many tropical regions have two seasons: the rainy, wet, or monsoon season and the dry season. Some people have a cool, mild, or harmattan third season. Seasons can also be determined by the timing of major biological phenomena such as hurricane season, tornado season, and wildfire season. [citation needed] Historical significance can be found in the ancient Egyptian seasons of flood, growth, and low water, which were historically defined by the annual flooding of the Nile in Egypt.

4) Soil type

Soil is made up of organic matter, minerals, gases, liquids, and organisms, all of which work together to support life. The pedosphere, or body of soil on Earth, has four purposes:

- as a substrate for plant development
- as a source of water storage, supply and purification
- as a moderator of Earth's atmosphere
- as a habitat for species

Soil is also known as earth or dirt; some scientific definitions distinguish soil from dirt by limiting the former term to displaced soil only. The pedosphere interacts with the lithosphere, hydrosphere, atmosphere, and biosphere.
The term pedolith, which is most commonly used to refer to soil, means "ground stone" in the sense of "basic stone" and is derived from the ancient Greek v, which means "ground, earth." Soil consists of two phases: a solid phase composed of minerals and organic matter (the soil matrix) and a porous phase composed of gases (the soil atmosphere) and water (the soil water) (the soil solution). Soil scientists can think of soils as a three-state system consisting of solids, liquids, and gases.

The term pedolith, which is most commonly used to refer to soil, means "ground stone" in the sense of "basic stone" and is derived from the ancient Greek v, which means "ground, earth." Soil consists of two phases: a solid phase composed of minerals

and organic matter (the soil matrix) and a porous phase composed of gases (the soil atmosphere) and water (the soil water) (the soil solution).

Soil is the result of multiple variables interacting over time, including climate, relief (elevation, orientation, and slope of terrain), organisms, and the soil's parent materials (initial minerals). It develops throughout time as a result of a variety of physical, chemical, and biological processes, including weathering and erosion. Soil ecologists consider soil to be an ecosystem because of its complexity and strong internal connections.

IV. RESULTS AND DISCUSSION

This proposed model is divided into two phases: Prediction and Recommendation. In prediction phase it takes all parameters as input and gives the best category of plant that can be grown in those conditions as output.

CONFUSION MATRIX FOR BINARY CLASSIFICATION

The confusion matrix	**Predicted**	**Predicted**
Actual	TP	FN
Actual	FP	TN

Sensitivity:
The chance of a positive test, conditioned on it being positive, is known as sensitivity (true positive rate).

$$Sensitivity = \frac{TruePostive\,(TP)}{TruePostive\,(TP) + FalseNegative\,(FN)} \quad (1)$$

Specificity:
The chance of a negative test if it is negative is referred to as specificity (true negative rate).

$$Specificity = \frac{TrueNegative\,(TN)}{True\,Negative\,(TP) + FalsePositive\,(FP)} \quad (2)$$

Precision:
Precision is the amount of information provided by a number in terms of its digits; it indicates how closely two or more measurements are related. It is indifferent by accuracy.

$$Precision = \frac{TruePostive\,(TP)}{TruePostive\,(TP) + FalsePositive\,(FP)} \quad (3)$$

Accuracy:

The measure of performance is to use accuracy. The accuracy is the percentage of correctly identified cases (both genuine and fake users) in the total number of examined cases, which can be calculated using the equation

$$Accuracy = \frac{TruePostive\,(TP) + TrueNegative(TN)}{TruePostive\,(TP) + FalsePostive\,(FP) + FalseNegative\,(FN) + TrueNegative(TN)} \quad (4)$$

A. Results Comparison

The below bar plot shows the comparison made among different machine learning algorithms. On x-axis four different machine learning algorithms are considered. On y-axis accuracy is taken. The classifiers that are considered are Decision Tree classifier, Random Forest classifier, K-nn classifier and Svm classifier. From all these classifiers, the classifier which gave the majority voting or highest accuracy with better is considered.

From the below shown Fig.3, it is observed that all the remaining three classifiers works better with the model except the Svm classifier which got least accuracy. Whereas, K-nn works much better than Svm. Decision Tree and Random Forest classifiers gave almost same results sightly differs in the decimals for the accuracy. Other measures like Precision, Recall and F1 score was also calculated for all the four machine learning classifiers.

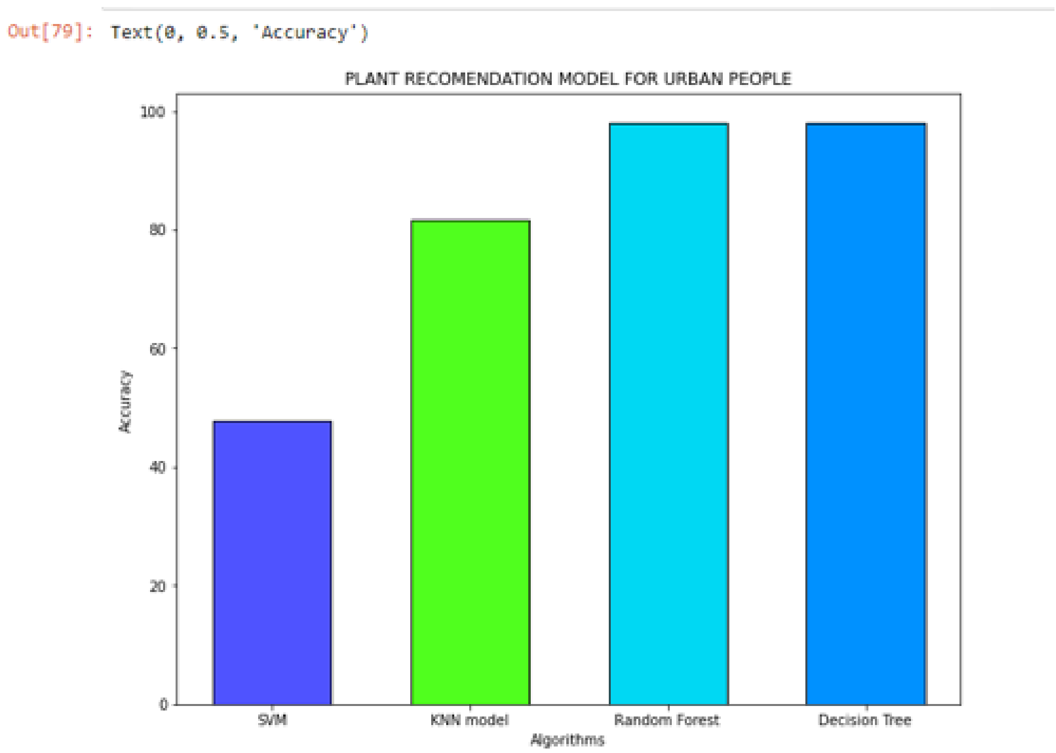

Figure 3. Results in comparison of the proposed method with previous methods

TABLE II EVALUATION METRIC VALUES COMPARISON

MODEL	PRECISION	F1 SCORE	RECALL	ACCURACY
Decision tree	0.96	0.96	0.97	96.90
Random forest	0.96	0.96	0.96	96.60
KNN	0.81	0.81	0.83	84.60
SVM	0.58	0.55	0.64	47.7

B. Accuracy

When compared with other machine learning models such as Random Forest, KNN and SVM which got the accuracies like 96.6%, 84.6% and 47.7%. Our model got the accuracy of 96.9% by using Decision Tree algorithm. Thus, we can say that this model works better using Decision Tree algorithm.

```
In [19]: from sklearn import metrics
         preds=model.predict(inputs_test)
         accuracy = metrics.accuracy_score(target_test, preds)
         accuracy

Out[19]: 0.9694656488549618
```

Figure 4. Accuracy of the model

V. Conclusion And Future Enhancement

This project Plant Recommendation Model for Urban People Using Majority Voting Classifier is cost-effective, eco-friendly, practical for recommending plants. Four types of classifiers are compared, namely Decision Tree, Support Vector Machine, Random Forest and K-Nearest Neighbor are implemented in this project. This model is able to predict the best category of plant and gives different recommendations to grow the plant based on the input. This model obtained the accuracy of 97% when tested with Decision Tree classifier and gives the best results when compared with other classifiers. The datasets considered in this project is our own customised dataset which contains important parameters to grow the plant such as temperature, humidity, season and soil type. The Plant Recommendation Model is useful people who do not have any kind of knowledge to grow the plant and also recommends the user with the best category of plant to grow.

In future, this project can be extended using a large data set. The dataset currently used for implementation includes 1300 records, which can be increased to get better results. Additional parameter such as user's locality can be added. If the growth of the plant is not up to the mark then soil testing can be included to provide high end results.

References

[1]. M Michael Rzanny, Patrick Mader, Alice Deggelmann, Minqian Chen, Jana Waldchen, "Flowers, leaves or both? How to obtain suitable images for automated plant identification", Plant Methods, 2019.

[2]. Israa Mohammed Hasson, Samar AmilKassir, Shyma Mohammed Altaie, "A Review of Plant Species Identification Techniques", International Journal of Science and Research (IJSR), 2018.

[3]. Jana Waldchen, Michael Rzanny, Marco Seeland, Patrick Mader, "Automated plant species identification – Trends and future directions", Plos Computational Biology, 2018.

[4]. Luciano D.S. Pacifico, ValmirMacario, Joao F.L. Oliveria, "Plant Classification Using Artificial Neural Networks", IEEE, 2018.

[5]. Neha Goyal, Kapil, Nitin Kumar, "Plant Species Identification using Leaf Image Retrieval", IEEE, 2018.

[6]. Parul Mittal, Manie Kansal, Hardeep Kaur, "Combined Classifier for Plant Classification and Identification from Leaf Image based on Visual Attributes", IEEE, 2018.

[7]. S. Anubha Pearline, V. Sathiesh Kumar, S. Harini, "A study on plant recognition using conventional image processing and deep learning approaches", Journal of Intelligent and Fuzzy Systems, 2018.

[8]. E M Imah, Y S Rahayu, A Wintarti, "Plant Leaf Recognition Using Competitive Based Learning Algorithm", IOP Conference Series: Materials Science and Engineering, 2018.

[9]. Cixiao Wang, "The Research on the Application of Plant Identification and Mobile Learning App based on Expert System", Science and Technology Publications, 2017.

[10]. Yu Sun, Yuan Liu, Guan Wang, Haiyan Zhang, "Deep learning for Plant Identification in Natural Environment", Hindawi, 2017.

[11]. Amala Sabu, Sreekumar K, Rahul R Nair, "Recognition of Ayurvedic Medicinal Plants from Leaves: A Computer Vision Approach", IEEE, 2017.

[12]. Jana Waldchen, Patrick Mader, "Plant Species Identification Using Computer Vision Techniques: A Systematic Literature Review", Springer, 2017.

[13]. Aparajita Sahay, Min Chen, "Leaf Analysis for Plant Recognition", IEEE, 2016.

[14]. TaisongJin, Xueliang Hou, Pifan Li, Feifei Zhou, "A Novel Method of Automatic Plant Species Identification Using Sparse Representation of Leaf Tooth Features", Plos One, 2015.