PLANT RECOMMENDATION MODEL FOR URBAN PEOPLE USING MAJORITY VOTING CLASSIFIER

Project Report

Submitted in Partial Fulfillment of the Requirements for the Award of Degree of

Bachelor of Technology

In

COMPUTER SCIENCE AND ENGINEERING

By

V. DIVYA PRIYANKA	18K61A05G7
S. NOOKAMBIKA	18K61A05D4
SK. IMRAN FAZIL	18K61A05E0
V. SUMANTH	18K61A05H0

Under the esteemed guidance of

Mrs. M. ANANTHA LAKSHMI M. Tech
Assistant Professor, CSE

Department of Computer Science and Engineering

SASI INSTITUTE OF TECHNOLOGY&ENGINEERING

(Approved by AICTE, New Delhi, Permanently Affiliated to JNTUK, Kakinada, and SBTET-Hyderabad, accredited by NAAC with ‘A’ Grade and NBA, ranked as "A" Grade by Govt. of A.P., Recognised by UGC 2(f) & 12(B))

Kadakatla, TADEPALLIGUDEM– 534 101

ACADEMIC YEAR 2021-2022

VISION AND MISSION OF INSTITUTE

VISION

Confect is a premier institute for professional education by creating technocrats who can address society's needs through inventions and innovations.

MISSION

1. Partake in the national growth of the technological, industrial, industrial area with societal responsibilities.
2. Provide an environment that promotes productive research.
3. Meet stakeholder's expectations through continued and sustained quality improvements.

VISION AND MISSION OF DEPARTMENT

VISION

Build a learning environment that enhances creativity, social awareness, and leadership skills for total personality development.

MISSION

The Computer Science and Engineering department's consistent effort is to provide the learner with exposure to emerging technology by providing hands-on experience making them creative and research-oriented professionals with values, Leadership qualities, and zeal to serve society.

PROGRAM OUTCOMES (POs)

Students in the Computer Science and Engineering program should, at the time of their graduation be in possession of:

PO1.Engineering Knowledge:

Apply knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

PO2.Problem Analysis:

Identity, formulates, research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO3.Design/ Development of Solutions:

Design solutions for complex engineering problems and design system components or processes that meet specified needs with appropriate consideration for public health and safety, cultural, societal, and environmental considerations.

PO4. Conduct investigations of complex problems:

Using research-based knowledge and research methods including design of experiments, analysis, and interpretation of data, and synthesis of the information to provide valid conclusions.

PO5.Modern Tool Usage:

Create, select and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

PO6.The Engineer and Society:

Apply to reason informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to professional engineering practice.

PO7.Environment and Sustainability:

Understand the impact of professional engineering solutions in societal and environmental contexts and demonstrate knowledge of and need for sustainable development.

PO8.Ethics:

Apply ethical principles and commit to professional ethics and responsibilities and norms of engineering practice.

PO9. Individual and Team Work:

Function effectively as an individual, and as a member or leader in diverse teams and multidisciplinary settings.

PO10.Communication:

Communicate effectively on complex engineering activities with the engineering community and with society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations and give and receive clear instructions.

PO11.Life-long Learning:

Recognize the need for and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

PO12.Project Management and Finance:

Demonstrate knowledge and understanding of engineering and management principles and apply these to one's work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

PROGRAM SPECIFIC OUTCOMES (PSOs)

PSO1.Mobile & Web Application Development:

Ability to develop mobile & web applications using J2EE, Android, and J2ME.

PSO2.Cloud Services:

To deploy virtualized and cloud-based services in the organization.

PROGRAM EDUCATIONAL OBJECTIVES (PEOs)

PEO1: Graduates will be able to analyse, design, and develop advanced computer applications to provide solutions to complex problems.

PEO2: Graduates are well-trained, confident, research-oriented and industry-ready professionals who are intellectual, ethical, and socially committed.

PEO3: Graduates will have the technical, communication skills and character that will prepare them for technical and leadership roles.

COURSE OUTCOMES (COs)

CO1. Develop problem formation and design skills for engineering and real-world problems.
CO2. Collect and Generate ideas through literature surveys on current research areas which help to analyse and present to impart knowledge in different fields.
CO3. Impart knowledge on software & hardware to meet industry perspective needs and standards.
CO4. Create interest to research innovative ideas as lifelong learning.
CO5. Ability to work with a team, and enrich presentation and communication skills.
CO6. Create a platform that makes students employable.

EXPECTED OUTCOMES

PROGRAM OUTCOMES (POs)

PO1: Engineering Knowledge

PO2: Problem Analysis

PO3: Design/Development of Solutions

PO4: Conduct investigation of complex problems

PO5: Modern Tool Usage

PO6: The Engineer and Society

PO7: Environment and Sustainability

PO8: Ethics

PO9: Individual Team Work

PO10: Communication

PO 11: Life-long Learning

PO12: Project Management and Finance

PROGRAM SPECIFIC OUTCOME (PSOs)

PSO1: **Mobile & Web Application Development**

PSO2: **Cloud Services**

SASI INSTITUTE OF TECHNOLOGY&ENGINEERING

(Approved by AICTE, New Delhi, Affiliated to JNTUK, Kakinada, and SBTET-Hyderabad Accredited by NAAC with 'A' Grade and NBA) Kadakatla, TADEPALLIGUDEM – 534 101, WG-Dist, A P, Ph: 08818 244989

Department of Computer Science and Engineering

sasi autonomous INSTITUTE OF TECHNOLOGY & ENGINEERING

Accredited by NAAC with "A" Grade, Recognised by UGC under section 2(f) &12(B) Approved by AICTE - New Delhi Permanently Affiliated to JNTUK, SBTET, Ranked as "A" Grade by Govt. of A.P.,

CERTIFICATE

*This is to certify that the project work entitled "**Plant Recommendation Model For Urban People**" is being submitted* **by** ***V.DIVYA PRIYANKA (18K61A05G7), S.NOOKAMBIKA (18K61A05D4), SK. IMRAN FAZIL (18K61A05E0), V.SUMANTH (18K61A05H0)*** *in partial fulfillment for the award of the degree of* ***BACHELOR OF TECHNOLOGY****, in* ***Computer Science and Engineering*** *to Jawaharlal Nehru Technological University, Kakinada during the academic year 2021 to 2022 is a record of Bonafede work carried out by them under my/our guidance and supervision. The results presented in this thesis have been verified and are found to be satisfactory. The results embodied in this thesis have not been submitted to any other University or Institute for the award of any other degree or diploma.*

Project Supervisor	**Head of Department**
Mrs. M. ANANTHA LAKSHMI	**Dr.M.V.S.S.NAGENDRANATH**
Assistant Professor	**Professor & HOD**
Department of CSE	**Department of CSE**

External Examiner

SASI INSTITUTE OF TECHNOLOGY & ENGINEERING (autonomous)

Accredited by NAAC with "A" Grade,
Recognised by UGC under section 2(f) &12(B)
Approved by AICTE - New Delhi
Permanently Affiliated to JNTUK, SBTET,
Ranked as "A" Grade by Govt. of A.P.,

DECLARATION BY THE CANDIDATES

We **V. Divya Priyanka (18K61A05G7)**, **S. Nookambika(18K61A05D4)**, **Sk. Imran Fazil (18K61A05E0), V.Sumanth (18K61A05H0),** hereby declare that the project report entitled "**Plant Recommendation Model for Urban People**" carried out under esteemed supervision of **Mrs. M. Anantha Lakshmi**, is submitted in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science & Engineering. This is a record of work carried out by us and the results embodied in this project has not been reproduced or copied from any source. The results embodied in this project report have not been submitted to any other University or Institute for the award of any other degree or diploma.

V.Divya Priyanka	18K61A05G7
S. Nookambika	18K61A05D4
Sk. Imram Fazil	18K61A05E0
V. Sumanth	18K61A05H0

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 a decision tree algorithm 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, Recommender Systems, Decision Tree, Plant Recommendation.

ACKNOWLEDGEMENT

We pay obeisance to our dynamic Chairman, **Sri B. Venu Gopala Krishna**, Sasi Educational Society, for his inspiring presence, which has always been the principle driving force behind all over endeavours.First of all, we would like to extend a special thanks to **Sri. M. Narendra Krishna**, Vice Chairman, Sasi Institute of Technology and Engineering, for his everlasting support.

It is a matter of great honour to thank **Dr. K. Bhanu Prasad**, Director, Sasi Institute of Technology and Engineering, who is behind us to support in all the aspects to complete the project.It is true pleasure for us to thank **Dr. J. Anand Chandu Lal**, Principal, Sasi Institute of Technology and Engineering, who is the striving force for us to make this project through periodical audits.

We feel privilege to thank **Dr. M. V. S. S. Nagendranath**, Profesor and Head of the Department, Computer Science and Engineering, for providing us with invaluable feedback on our work, which allowed us to constantly improve our project.

We are very much grateful to thank our supervisor **Mrs. M.Anantha Lakshmi**, Assistant Professor, Sasi Institute of Technology and Engineering for her constant encouragement, monitoring, guidance and live with us throughout the submission of the project. She motivated us whenever we encountered an obstacle along the way.

We are very happy to thank our project coordinator **Dr. P. Sivakumar** and to all the Project Evaluation Committee members who provide valuable suggestion to fine tune our ideas and projects. We are also thankful to all teaching and non-teaching staff members those who contributed well for the successful completion of our project work.

With gratitude,

V. Divya Priyanka 18K61A05G7

S. Nookambika 18K61A05D4

SK. Imran Fazil 18K61A05E0

V. Sumanth 18K61A05H0

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

ABBREVIATION:

AES	-	Advanced Encryption Standard
CNN	-	Convolution Neural Networks
SVM	-	Support Vector Machine
K-NN	-	K-Nearest Neighbour
LDA	-	Linear Discrimination Analysis
RF	-	Random Forests
SIFT	-	Scale Invariant Feature Transform
PNN	-	Probabilistic Neural Network
GRNN	-	General Regression Neural Network
MMC	-	Move Median Centres
MLP	-	Multi-Layer-Perceptron
DT	-	Decision Tree
RBPNN	-	Radial Basis Probabilistic Neural Network
CPSRS	-	Computerized Plant Species Recognition System
APSR	-	Automatic Plant Species Recognition
MCH	-	Moving Centre Hypersphere
APIS	-	Advanced Plant Identification System
RBPNN	-	Radial Basis Probabilistic Neural Network
RBFNN	-	Radial Basis Function Neural Networks

CHAPTER 1

INTRODUCTION

1.1 PREAMBLE

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.

1.2 OVERVIEW OF THE PROJECT

The overview of the project contains the following details:

1.2.1 Need For Plants

Plants have a major role on Earth, producing essential resources such as oxygen while also maintaining the ecological balance. The relationship between plants and humans is a long and continuous one. Plants are used in both practical and symbolic ways by humans, such as for food, clothing, furniture, medicine, and, much more from them. Plants have long been used as a kind of traditional medicine. It is considered low-cost and has no negative side effects. Plants are also used as a feedstock for a variety of industrial products, including timber and paper, as well as a variety of chemicals. They are also used in art, mythology, and literature. Plants provide enjoyment to millions of people through gardening. Although living in a society, all have not lost this dependence on plants.

1.2.2 Parts of a plant

Roots, stems, flowers, fruits, leaves, and seeds are the basic elements of most terrestrial plants. The function of each plant part is described below.

- **Root**

Roots hold plants in the soil and absorb nutrients and water for the remainder of the plant. Dermal, ground and vascular tissues are all present in each root. The roots are the organs of a plant that are modified to provide anchorage for the plant and take in water and nutrients into the plant body, which allows plants to grow taller and faster.

- **Stem**

Stems sustain the plant's top section and serve as a channel for nutrients, water, sugar, and carbohydrates. Some plants, such as cactus, celery, asparagus, and bananas, can undergo photosynthesis in their stems.

- **Flower**

Flowers are a plant's reproductive organ. To attract pollinators such as birds, bees, and other insects, they generally feature spectacular petals and perfumes. The four primary elements of most flowers are petals, stamen, pistil, and sepals. Flowers generate seeds in their ovary after they have been pollinated and fertilized.

- **Fruit**

Fruits are the fleshy materials that enclose seeds. They serve to protect the seeds while also attracting animals to devour them. This aids in the dispersal of seeds. They serve to protect the seeds while also attracting animals to devour them. This aids in the dispersal of seeds. They serve to protect the seeds while also attracting animals to devour them.

- **Leaves**

  Photosynthesis, or the production of food for the plant, usually takes place in the leaves. Chlorophyll is a green molecule that collects light energy and converts water and carbon dioxide into plant food and oxygen.

- **Seed**

Seeds are made up of plant material that can grow into a new plant. An embryo is a name given to this plant substance. Seeds have one or two cotyledons and are protected by a seed coat. Cotyledons are the young plant's food until it can make its food from light, and they are often the plant's first embryonic leaves.

1.2.3 Division of plant

The plant is also divided into monocarpic and polycarpic plants. Monocarpic plants only blossom and reproduce once their life and then die. It comes in two varieties: annual and biennial. Annual plants, such as sunflower, rice, and wheat, grow once a year, but biennial plants, such as radish, spinach, and bottle gourd, grow twice a year. Polycarpic plants are perennials that flower every year at a specific season after reaching maturity.

1.2.4 Species of a plants

A species is defined as a group of animals, plants, or other living things that have similar features and are categorized as similar in some way. All over the world, there are currently about 310000–420000 known plant species, and many are still unknown yet. Approximately 7,000 plant species have been utilized for food, with approximately 30 species accounting for the majority of today's food. Cereals like rice and wheat, starchy roots and tubers like cassava and potato, and legumes like peas and beans are all important mainstays. Lipids are provided by vegetable oils such as olive oil, whereas vitamins and minerals are provided by fruits and vegetables. Plant recognition or classification can be done based on parts of plant-like: Leaves, flowers, and fruits of the plant. Cereals like rice and wheat, starchy roots and tubers like cassava and potato, and legumes like peas and beans are all important mainstays. Lipids are provided by vegetable oils such as olive oil, whereas vitamins and minerals are provided by fruits and vegetables. Cereals like rice and wheat, starchy roots tubers like cassava and potato, and legumes like peas and beans.

1.2.5 Season based plants

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.

1.3 PROBLEM STATEMENT

Knowing plant species is very important in today's digital world as plants are essential in our life. Numerous studies have investigated various methods and models to identify species of plants.

However, these methods have traditionally focused on increasing the accuracy of prediction rather than taking more datasets during the training of the model which leads to error-prone and decreasing the performance of the models. As a result, the existing research is inadequate for season-based plant classification.

1.4 AIM OF THE PROJECT

- There are no classifications made for plants based on the season in the literature.
- This method or model aims to identify plant species and classify them based on the season.

1.5 OBJECTIVE OF THE PROJECT

- **Objective 1:** To identify the species of the plant from the leaf.
- **Objective 2:** To classify the plant based on the season.
- **Objective 3:** To evaluate the performance of the proposed model by conducting rigorous experiments.
- **Objective 4:** To compare and contrast these models in terms of their accuracy and loss.

1.6 EXISTING SYSTEM

Plant classification or recognition can be done based on parts of plant-like: leaves, fruits, seeds, and flowers. Seeds, fruits, and flowers are not a good choice for plant species identification because they may not be present with the plant all the time. In the case of a few plants, they grow only in a particular season so that they won't produce flowers or fruits in other seasons due to this reason choosing leaves is a better option. Because leaves stay with the plant all the time. Even few plants have unique leaves which are easily identified.

In real practice, to begin identifying a plant, the user must match the plant specimen in hand, with the precompiled plant characteristics descriptions prepared by the plant taxonomist. This traditional method of plant identification is complex and time-consuming. However, because the manual interpretation involves human visual perception, the outcome may not be precise.

1.6.1 Machine learning

Few machine learning algorithms are used for identifying plant species and comparisons are made among them. Machine learning algorithms are divided into two types supervised and unsupervised.

1.6.1a Supervised algorithms

The proposed multi-organ identification for several parts such as entire plant, flower frontal and lateral view, leaf top and back-side view and trained the model with Convolution Neural Networks (CNN), only flower front, flower lateral and leaf top view gave more accurate results but was only considered for few plants that are small dataset is considered [1]. And using image processing techniques and neural networks.

Image processing techniques are used to remove the noise of the image and the Image extraction process is consist of RGB color extraction, leaf shape extraction, and leaf vein pattern extraction but it suffers from the practical issues of relying on the use of quality cameras [18].

The proposed two approaches traditional method and the deep learning methodology are the two approaches. Hu moments (shape characteristics), Hara-lick texture, local binary pattern (texture features), and color channel statistics are used in the classic method to extract features (color features).

Different classifiers, such as linear discriminant analysis, logistic regression, classification and regression tree, naive Bayes, k-nearest neighbor, random forest, and bagging classifier, are used to classify the retrieved features. Using dataset augmentation methodologies, the accuracies can be enhanced by increasing the number of images [7].

The proposed mobile app allows users to identify plants from photographs of their leaves. Pre-processing, feature extraction, and feature matching are the three components. A weighted k-nearest neighbor search algorithm is used to identify and return the matched plant species but only fewer species are taken into consideration [13]. A comparison of Linear Discrimination Analysis (LDA) and Random Forests (RF) classifiers has been presented. Pre-processing, feature extraction, and plant classification are the three key stages of the proposed system. And showed that LDA achieved the highest classification accuracy compared to RF. To improve results, the classification of different types of plants can be done by involving other features [17].

Picture acquisition, image pre-processing, feature extraction, identification, and performance measures were the five steps of the previous study. The phase of feature extraction is the most critical and crucial in plant identification. To identify plants, the Scale Invariant Feature Transform (SIFT) method of shape feature extraction and the Grid-Based Colour Moment (GBCM) approach of color feature extraction were discussed. But more robust features such as texture and also intend to experiment with learning classifiers such as supervised classifiers to increase the identification rate has to be included [20]. The previously proposed methodology, combined image and data processing techniques with Probabilistic Neural Network (PNN). PCA is used to orthogonalize 12 leaf features that are extracted from five fundamental features: diameter, physical strength, physiological width, leaf area, and leaf perimeter but have taken very less plants that can be classified [26].

The previously proposed method focused on what kinds of leaf features should be extracted, what external aspects should be addressed before starting the extraction process, and what kinds of extraction and classification. algorithms can be applied for plant recognition and classification. Comparison of different classifiers like Probabilistic Neural Network (PNN), Probabilistic Neural Network with color and texture, Moving Centre Hypersphere Classifier and k- Nearest Neighbors, Euclidean Distances, Linear Discriminant Analysis (LDA) with Nearest Neighbor (1-NN) and General Regression Neural Network (GRNN) were discussed. And selected three classifiers PNN with color and texture, LDA with Nearest Neighbor (1-NN), and General Regression Neural Network (GRNN) these classification algorithms gave the highest accurate results but this has more mathematical computations [19].

The proposed plant image recognition approach is based on digital morphological features to achieve classification, a new classification method known as the move median centers (MMC) hypersphere classifier is used. When compared to the 1-NN and k-NN classifiers, it is discovered that the MMC classifier not only saves storage space but also reduces classification time without reducing classification accuracy but it considered only a fewplants[27].

A Multi-Layer-Perceptron (MLP) artificial neural network trained with a Back propagation algorithm to perform automatic plant classification. Comparison of five different classifiers from plant classification literature and some of their variants: Naive Bayes, Decision Tree classifier (DT) K-Nearest Neighbors (KNN, with k = 3, 4, and 5) classifier, Support Vector Machine with k = 3, 4, and 5 A Multi-Layer Perceptron with RBF (SVMrbf) and Linear (S V M linear) kernel functions, as well as an RBF (SVMrbf) and Linear (S V M linear) kernel function with the algorithm of backpropagation (MLP-BP) but was only considered small dataset [4]. 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 but this model has more complexity and difficulty [8].

A novel framework for recognizing and identifying plants that combine Zernike movements with shape, vein, color, and texture traits. As a classifier, a radial basis probabilistic neural network (RBPNN) was used but the negative is Zernike moments is costly computation, that is this model needs more time for calculations [21]. Only plants with broad flat leaves that are more or less two-dimensional operate with this strategy.

Eccentricity, area, perimeter, major axis, minor axis, equivalent diameter, convex area, and extension are among the morphological traits recovered. These characteristics are fed into the probabilistic neural network as inputs. A ten-fold cross-validation methodology was used to test the suggested strategy. This system's most significant flaw is that it requires user assistance during the pre-processing stage. Another drawback is that it is unable to work with images with complex backgrounds [23]. This research introduced a new wavelet decomposition and Gaussian interpolation method that successfully isolate not only the contour but also the venation, or the leaf's structure. Different classifiers, such as 1-NN, k-NN, and RBPNN, used run-length features (RF) to achieve leaf recognition. It has a higher rate of proper recognition when employing features taken from the skeleton than other approaches but this model has more mathematical computations [29].

Decision tree classifiers in the experiments but fails at the dataset in this study is based on a white backdrop, advancement of the algorithm can be achieved to detect category of leaf picture from the colored background as well as in multiple object image [6].

The performance of two widely used classifiers: Support Vector Machine (SVM) and Probabilistic Neural Network (PNN) is examined in terms of Precision, Recall, and F-Score and was discussed.

1.6.1b Computer vision techniques

A Computerized Plant Species Recognition System (CPSRS) is a web-based application with two parts: a web client and a web server were built. CPSRS has two main functions: information storage and retrieval, as well as automatic picture and text processing. Text-based information retrieval and content-based leaf retrieval are two types of plant species retrieval methods explored. If the focus on the methods of Automatic plant species recognition (APSR) is made then the recognition accuracy of the input image will be increased [30].

A system, whose key aspects rely on computer vision was proposed which includes determining whether images are 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 neighbor’s search. Color-based segmentation outperforms other published segmentation algorithms, due to the task's restrictions, which include preserving fine-scale features such as thin structures and operating at interactive speeds. But collaborative features can be added to the app i.e., for entire classrooms to use together [22]. A model using computer vision to recognize ayurvedic medicinal plant species located in India's the Western Ghats was recommended. It employs a combination of SURF and HOG features taken from leaf images, as well as a K-Nearest Neighbor (KNN) classifier for classification but this model is useful only for plant species in that specific area [11].

Different techniques of plant identification and feature extraction methods were discussed. And also discussed different Identifying Plants. There are two types of approaches. Approaches based on leaves: Aspect ratio, narrow factor, compactness, centroid, eccentricity, dispersion, area, equivalent diameter, moments invariant, and other essential leaf properties are extracted in most research based on leaf shape or color. Approaches are based on the area, perimeter, roundness, and aspect ratio of flowers or fruits in this flower shape descriptor but to improve the accuracy chemical methods must be used to treat the leaves [2]. The goal of this paper was to identify, analyze, and compare research on plant species identification using computer vision techniques.

1.7 METHODOLOGY:

Although many approaches have been proposed and tested with almost the entire leaf features successfully extracted and recognized, still those approaches have their limitations. Some approaches are found to be inaccurate, primarily because the input image contains noise. Besides that, different understanding of consideration of the extracted features also influence the finding because a different definition of the features or different dataset has been used for testing.

1.7.1 Data collection

Seasonally, consider the limited flowering, fruiting, veggie, and herb plants. After that, train the data set and test the algorithm by taking records from the dataset. The data plays an eminent role in training the algorithm. The bigger the data the bigger the complexity involved in it. Here the current data set contains a few parameters such as season, temperature, humidity, and soil type. These are the important parameters required to grow a plant.

1.7.2 Classification of plants

These three traits are used to classify plants:

- Evergreen plants are those that keep their leaves throughout the year (all year round).
- Woody plants are classified as either deciduous or evergreen.

- Deciduous plants are those that shed their leaves at the end of the growing season, either in the winter in temperate climates or during the dry season in tropical climates.

The life cycle of plants can be used to classify them.

1.7.2a Annuals

These are plants that go through their entire life cycle in just one season. Normally, they are herbaceous. Annual plants include corn, rice, wheat, and legumes, to name a few.

1.7.2b Biennials

Plants that take two years to complete their life cycle are known as two-year plants. Normally, they are herbaceous. Carrots, cabbage, onions, and beets, for example, are biennial plants.

1.7.2c Perennials

These are plants that live for an extended period, usually more than two years. They have a woody or herbaceous appearance. Perennials include roses, lavender, dianthus, and lilies, to name a few.

1.7.3 Season based plant classification

The time of year when the weather and temperature allow plants to develop is known as the growing season. A growth season might be as brief as four months in some locations. A growing season can span over a year in more tropical and warm climates. Summer, fall, winter, and spring are not to be confused with the seasons known as summer, fall, winter, and spring

1.7.3a Basic conditions for season-based plants

The best circumstances are determined totally by the needs of the plant. Each type of plant has its own set of demands and requirements, although light, carbon, hydrogen, and oxygen are all important. Here is a quick rundown of some of the environmental factors that affect plant growth:

- Day-length

The most important component in regulating vegetative growth, flower initiation and development, and dormancy induction is usually the length of the day.

Plants use the length of the day as a cue to boost growth in the spring and prepare for the winter. To flower, many plants require precise day-length conditions.

- Light

Light is the energy source for plants. Cloudy, rainy days or the shade cast by neighboring plants and structures can greatly diminish the quantity of light accessible. Plants that have acclimated to the shade are unable to resist the harsh light of the full sun. Plants can only live if the amount is within their tolerance range. If the limit exceeds then the plant may grow Improperly which is intolerable.

- Temperature

Plants thrive in a specific temperature range, which might be large for some species and narrow for others. Plants can only exist in environments where they can carry out life-sustaining chemical reactions. Excess heat and insufficient heat or temperature can also have a large impact on the growth of the plants, ultimately this result won't be satisfactory if any lack of heat.

- pH

The ability of plant roots to absorb certain nutrients is influenced by pH, which is a measurement of your soil& acidity or alkalinity. Most plants thrive in soils with a pH of around 7.0. To thrive, most ericaceous plants, such as azaleas and blueberries, require acid soils with a pH below 6.0. Materials containing sulphates, such as aluminium sulphate and iron sulphate, can be used to raise pH and materials containing sulphates, such as calcium ulphate, can be used to lower pH.

- Water

Plants require not just an acceptable amount of water, but also the correct type of water. A nutritional imbalance and poor plant growth can occur when water contains too much alkalinity or salt. Water is one of the most important minerals which play a major role in photosynthesis process where a plant prepares its own food using other factors. Water is one of the most important minerals which play a major role in photosynthesis process where a plant prepares its own food using other factors. A nutritional imbalance and poor plant growth can occur when water contains too much alkalinity or salt.

- Oxygen

Plants require oxygen for respiration in order to carry out their water and nutrient absorbing functions. Oxygen helps organisms grow, reproduce, and turn food into energy. Plant cells are constantly breathing. Plants make their own oxygen when their leaves are lit.

- Nutrients

Macronutrients and micronutrients are the two types of nutrients found in plants. Macronutrients are the essential nutrients that plants use to survive. Minerals, carbon, hydrogen, and oxygen are among them. Micronutrients are secondary mineral elements that are used in smaller amounts than primary nutrients. The 12 most common necessary mineral nutrients are listed below. These minerals are provided in the garden by the soil as well as fertilizers such as manure, compost, and fertilizer salts. Nitrogen, phosphorus, potassium, calcium, magnesium, and sulphur are all important elements in considerable proportions. Iron, manganese, boron, zinc, copper, molybdenum, and chlorine are also required, albeit in very small quantities.

1.8 SIGNIFICANCE OF PROPOSED WORK

The primary goal of plant classification is to ensure that the correct plants are named, categorized, and identified. The ability to identify and collect plant species is utilized to safeguard your land from invasive species and to get the most out of it. There are names for all known plant species. Unfortunately, they do not come with name tags outside of flower shops and botanical gardens.

As a result, identifying an unknown plant, that is, determining the species to which it belongs and consequently its name, is sometimes required. The process of identifying plants implies that they have already been classed and identified. A dichotomous key is a stepwise elimination method that uses a sequence of paired contrasting statements.

1.8.1 Advantages of the proposed ideology

- Cheap in cost as there is no hardware cost involved in it.
- Builds accurate knowledge of plant identification and classification based on seasons.

- This kind of information is the knowledge basis for other fields, including pharmaceutical, biotechnology, and food industry.
- User friendly product any one with basic English reading skills.
- Effective in implementation for social use with higher accuracy

1.9 LIMITATIONS OF THE PROJECT

- The dataset that is considering is small to reduce complexity.
- The data is prior focused on the plants which can be grown in Andhra Pradesh.

1.10 ORGANIZATION OF THE REPORT

The rest of the chapter is organised as follow.

Chapter 1- This chapter contains an introduction, problem statement, aim of the project, methodology, significance of the work, and conclusion.

Chapter 2- This Chapter contains a literature review and comparison of various fake review detection techniques.

Chapter 3- This chapter contains System requirements like Hardware Tools, Communication interfaces, Software requirements/Hardware requirements.

Chapter 4- This chapter contains System design including System Architecture of the model.

Chapter 5- This chapter contains the implementation and proposed approach of the project.

Chapter 6- This chapter contains experimental results, project outcome and comparison with the existing approaches.

Chapter 7- This contains conclusion and future enhancement of the proposed project.

1.10 SUMMARY

This chapter includes a brief introduction to overview of plants and their identification n and also classification concepts. And this chapter deals with the identifying problem statement, estimating the objective of the project, a brief introduction about the methodology used in the project, the significance of the project, organization of the project which includes the output of every chapter introduction. The next chapter is the Literature review which discusses various journal papers to obtain the specific problem statement by analyzing all the relevant work and information mentioned in that reference paper to understand the present problem statement existing in that particular.

CHAPTER 2

LITERATURE SURVEY

2.1 PREAMBLE

The purpose of the literature survey is to obtain a clear understanding of the existing problem in the particular area of the domain. By clearly understanding all the previous development and their works will provide the best way to obtain the perfect problem statement existing in the present condition.

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

2.2 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.

Esraa Elhariri et al, (2014) [17] presented classification based on Random Forest (RF) and Linear Discrimination Analysis (LDA) algorithms for classifying different types of plant. This proposed system has three main stages: pre-processing, feature extraction and plant classification. Initially, the first stage is about the elimination of image background by performing morphological erosion on a binary mask using a 5x5 structure element, then multiplying the RGB image by the binary mask. The images are then transformed from RGB to HSV colour space, which is widely used in the field of colour vision and is similar to human colour perception categories.

In addition, to compute vein characteristics, each image is converted to a Grey-scale image. In the second stage, as most of the plants have leaves. Characteristics such as shape, colour, texture, and border distinguish leaves from one another. For plant species categorization, this technique uses a combination features vector that incorporates all of these characteristics. Finally, in the classification stage a various combination of extracted features and computed accuracy using each one of these combinations. This proposed approach applied two different algorithms for classification of plants type LDA and RF.

Random Forests (RF) is one of the most well-known classification and regression algorithms, with the ability to accurately categorize enormous datasets. An ensemble of decision trees is generated by the Random Forests method. Linear Discriminant Analysis (LDA) is a widely used data classification and dimensionality reduction technique. When the within-class frequencies are unequal, Linear Discriminant Analysis can readily manage the situation, and their performance has been evaluated using randomly produced test data. The main idea is to select a linear transformation that best distinguishes between classes, and then classify in transformed space using metrics like Euclidean distance. The inputs for any classifier are training dataset feature vectors and their corresponding classes, whereas the outputs are plant types of each image in the testing dataset.

Finally, this proposed classification approach was implemented by applying background removal, and extracting color components for each image.Then, feature extraction was applied to each pre-processed image, HSV color moments, shape, first order texture, GLCM and vein features are obtained as features vectors. And then LDA and RF algorithms are used. The Experimental results showed that LDA achieved the highest classification accuracy of 92.65 % when compared to RF. The advantage is that it gives quicker results as only a small dataset is considered and is simple to implement. The disadvantage of this methodology is the classification of different types of plants is made by involving only few features and this system is not reliable as accuracy is not great.

Thibaut Beghin et al. (2010) [24], proposed an algorithm to classify plant-based shape and texture classification based on the shape of the leaf. Firstly, two contour signatures are calculated based on the distance and angle of contour points from the leaf's centre. The dissimilarities between the graphs are determined using the Jeffrey distance, and this step is repeated for each leaf of the dataset.

The contour signature method, which is used to classify objects, produces excellent results. Second, to capture the differences in the macro-texture of the leaves, a classification involving the Sobel operator is used. The orientation and magnitude of the edge gradients are used to create a histogram. Finally, using probability density functions, a method is built that combines lobe differentiation, shaped-based, and texture-based methods. The incremental classification method is designed to get the most out of each method individually. In this work, a dataset of 18 species of leaves is considered. The experimental results showed that only 10 species out of 18 are successfully classified with the greater classification rate. The advantage of this model is it classifies leaves using a combination of relatively simple methods. It is a valid and promising approach but a smaller dataset is considered which contains only few plant species and is difficult to implement as intermediate mathematical knowledge is required.

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.

The fundamental idea of MCH is that each of the patterns in n-dimensional space can be represented using a series of n-hypersphere. In other words, the training process of the MCH classification method can be regarded as a data compression process. When the training process is done, MCH classifier is required to be able classify any given input data point. The perpendicular distances from the data point to the surface of all hypersphere s are selected as the classification criteria. Twenty types of practical plant leaves are correctly categorized, according to experimental data, with an average classification rate of up to 92.6 %. The advantage of this system is that it gives better accuracy rate and is tested on many datasets but to implement this system a great mathematical knowledge is required.

2.3 VALIDATION OF PLANT SPECIES

S. Anubha Pearline et al. (2018) [7], suggested a plant species recognition system utilizing two approaches: the standard method and the deep learning approach. Hu moments (shape features), Hara- lick texture, local binary pattern (texture features), and colour channel statistics are used in the classic method to extract features (colour features). Linear Discriminant Analysis, Logistic Regression, Classification and Regression Tree, naive Bayes, k-nearest neighbour, random forest, and bagging classifier are used to classify the retrieved features.

Different deep learning architectures, such as VGG 16, VGG 19, Inception-v3, and Inception- ResNetv2, that are CNN-based pre-trained models, are also tested in the context of plant species recognition. There are three conventional datasets (Folio, Swedish Leaf, and Flavia), as well as one real- time dataset (Leaf12). The implementation is done utilizing the Python programming language and the OpenCV package.When compared to previous methods, the deep learning model exhibited a greater accuracy of 97.14 %. The advantage of this methodology is that it is easy to perform as it is implemented in python but very few images are considered in the dataset.

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 Histogram of Oriented Gradients is a feature descriptor for object detection in computer vision and image processing. 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. The main advantage of this work is focused on the features that are computationally difficult to extract. But the leaf dataset considered is small and instead of using K-NN classifier, using SVM or ANN classifier gives better results.

Aparajita Sahay et al. (2016) [13], designed a mobile application that helps users to identify plants from photographs of their leaves. It consists of three components: pre-processing, feature extraction and feature matching. Firstly, pre-processing is conducted for noise reduction. A morphological technique named erode-dilation is used to the grey scale to eliminate noise. It erodes away the boundaries of the foreground object. The pre-processing step removes unwanted noise from the image and reduces the number of false positive regions. Secondly, feature extraction identifies representative features and computes scale invariant feature descriptors.

The scale invariant feature transform consists of some major stages of computation: Scale-space extrema detection, Key point localization, Orientation assignment and Key point descriptor. Finally, feature matching is done by matching the plant species that are identified and returned using a weighted k-nearest neighbor search algorithm. The dataset from LeafSnap[8] which consists of different combinations of plant species is used. Using this application, users can photograph a single leaf on a white coloured background and submit it as an input query image. Application will analyse features of the leaf and identify the plant species in real time. The advantage of this proposed methodology is that it is easy to implement and gives promising results but only works for an image with white coloured background.

W.H. Rankothge et al. (2013) [18] proposed an intelligent system. Advanced Plant Identification System (APIS) is an intelligent system that can identify tree species from images of their leaves, and it does it with greater accuracy and in less time. APIS is a plant recognition system that uses image processing and neural network techniques to identify plants. To remove image noise, normalize the leaf area, decrease the undesirable white backdrop, and scale the leaf image, image processing techniques are applied. RGB colour extraction, leaf shape extraction, and leaf vein pattern extraction are all part of the image extraction process. The APIS system employs the 2D-FFT approach to offer rotational invariant to the input image. If the APIS neural network delivers more than one work result image, the APIS system will go one step further and match colour (RGB values), width, and height.

APIS has a 95% accuracy rate and provides results in a couple of minutes. The main advantage of this system is that it takes less time to respond to the output. The limitation suffers from the practical challenges associated with relying on high-quality cameras. And performance of the APIS system can be improved by using the HSV colour model instead of the RGB colour model. The HSV colour model can reduce image contrast and brightness errors, and it is independent of hardware or lighting/weather conditions.

Mohamad Faizal Ab Jabal et al. (2013) [19] suggested the types of leaf features that need to be extracted, external elements that should be considered before the extraction process, and the several types of extraction and classification methods that can be employed for plant recognition and classification. Different classifiers like Probabilistic Neural Network (PNN), Probabilistic Neural Network with color and texture, Moving Center Hypersphere Classifier and k-Nearest Neighbors, Euclidean Distances, Linear Discriminant Analysis (LDA) with Nearest Neighbor (1-NN) and General Regression Neural Network (GRNN) are compared. And got the accuracies like 90.31%, 93.75%, 92.4%, 92%, 94.3% and 97.98%.The experimental result shows that PNN with color and texture, LDA with Nearest Neighbor (1-NN) and General Regression Neural Network (GRNN) classification algorithms gave highest accuracy.

A. H. Kulkarni et al. (2013) [21], implemented a plant recognition algorithm by using an easy way to extract features like shape, vein, color, texture features which are combined with Zernike movements. Radial basis probabilistic neural network (RBPNN) has been used as a classifier. The benefits of radial basis function neural networks (RBFNN) and probabilistic neural networks are combined in the radial basis probabilistic neural network (RBPNN) model (PNN). A dual stage training approach can be used to speed up the training of RBP neural networks. The parameters of the basis function, which must correspond to the RBP units, are determined in the second stage using an unsupervised learning technique that requires the solution of a series of nonlinear equations. Finally, estimation of the output connection weights, centers, and widths of the RBP units is required for RBP neural network training.

Finally, in this paper, the author came up with a plant recognition algorithm by using an easy way to extract features like shape, vein, color, texture features which are combined with Zernike movements which yields an accuracy rate of 93.82%. The main advantage of this system is classifying structurally complex images, Zernike moments for feature descriptors is a feasible alternative and they also can be computed in parallel as a result computational performance of computers increases.

The only disadvantage of Zernike moments is their high computational cost, which makes them unsuitable for various applications.

Stephen Gang Wu et al. (2007) [26], employed Probabilistic Neural Network (PNN) with image and data processing techniques to implement a general-purpose automated leaf recognition for plant classification. The data processing techniques used in this model are image processing that is converting RGB image to binary image, next boundary is enhanced and then features are extracted. PCA is used to orthogonalize 12 leaf features that are extracted from five fundamental features: diameter, physical strength, physiological width, leaf area, and leaf perimeter.

The purpose of PCA is to present the information of the original data as the linear combination of certain linear irrelevant variables. The PNN has been trained using 1800 leaves to classify 32 different types of plants with a 93.6 %.This method is an accurate artificial intelligence strategy that is fast in execution, efficient in recognition, and simple to construct when compared to other systems but only few images are considered in the dataset.

Ji-Xiang Du et al. (2007) [27], suggested a method for automatically recognizing plant images based on digital morphological features. The digital morphology features generally include geometric features (GF) and invariant moment features (MF). Aspect ratio, rectangularity, area ratio of convexity, perimeter ratio of convexity, sphericity, circularity, and eccentricity are some of the geometric properties. To perform classification, a new classification method known as the Move Median Centers (MMC) hypersphere classifier is used.

The MMC's central concept is that each pattern class is represented by a sequence of "Hypersphere s," which are considered as a collection of "points" in traditional techniques. The fifteen features are used to classify 20 species of plant leaves. The results of the experiments showed that the proposed approach is both effective and efficient. When compared to the 1-NN and k-NN classifiers, it is discovered that the MMC classifier may not only minimize storage space but also reduce classification time.

Xiao Gu et al, (2005) [29], proposed a new wavelet decomposition and Gaussian interpolation method, which not only successfully separates the contour but also separates the venation, i.e., the skeleton of the whole leaf. Wavelet transform is used to decompose leaf's images and produce new resolution images and a series of detailed images. These detailed images are removed at the effort of blurring the image, removing noise and detecting edges from image etc. Gaussian interpolation brings the vivid leaf's skeleton at different scales. All the details distributed over the images were focused on only a single image. The algorithm was separated into two main parts. Part 1 is the segmentation of the leaf's skeleton and Part 2 is the extraction of features. After getting the leaf's skeleton, texture features can be extracted from them to recognize the leaves.

Then, run-length features (RF) were used to perform the recognition of leaves by different classifiers such as 1-NN, k-NN and RBPNN.The experimental results found that this method can obviously segment the leaf's skeleton, and the correct recognition rate of using features extracted from the skeleton is more improved than other methods. This approach is easy to implement and understand but expert knowledge on transformations is required.

Yanhua Ye et al. (2004) [30], designed a Computerized Plant Species Recognition System (CPSRS). CPSRS is a web-based programme that allows users to search for and identify plant species in the field in a familiar and efficient manner. It consists of two components: web client and web server. Both of them were constructed on the Java web infrastructure. CPSRS plays two major roles: information storage and retrieval, and automatic image and text information processing.

Two types of plant species retrieval methods: Text-based information retrieval and content-based leaf retrieval are discussed. With the text-based information retrieval method, the exact information of plant species is retrieved from the database according to the input searching criteria. The content-based leaf retrieval provides a way to retrieve information by analysing the input leaf image, which is different from the text-based information retrieval. The model is trained with 130 leaves of the simple type. For the content-based leaf retrieval, experimental results show a recall rate of about 71.4%. But this model gives lower recognition accuracy of input image. The content-based leaf retrieval provides a way to retrieve information by analysing the input leaf image, which is different from the text-based information retrieval.

2.4 COMBINED PLANT CLASSIFICATION

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.

2.5 AUTOMATED PLANT IDENTIFICATION

Michael Rzanny et al. (2019) [1], proposed that multi organ identification would considerably improve recognition rates, especially for species that are difficult to distinguish for humans. They devised a method for taking images of flowering plants in order to produce observations. Each observation consists of five in-situ photos of the same individual taken from different angles (entire plant, flower frontal and lateral view, leaf top and back side view).The presence of plant species must be monitored and managed as part of conservation biology and sustainable development, but this requires specialist knowledge in terms of species identification. They compiled a dataset with 100 observations for each of the 101 species.

Using this dataset, we trained convolutional neural networks and used score level fusion to determine the prediction accuracy for each perspective individually and in combination. The most appropriate compromise in terms of acquisition effort and accuracy is to combine flower frontal, flower lateral, and leaf top views (96 %).Finally, as a result, multi-organ observations, including at least the front and later a perspective of flowers and the leaf top view, would enrich herbaceous plant picture collections. The disadvantage is it took less data to identify multi – organ observations.

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.

Nursuriati Jamil et al. (2015) [15], proposed A study is being done to determine which of three low- level traits for plant leaf identification contributes the most. 455 herbal medicinal plant leaves were used for intra-class and inter-class identification, with 70% of the dataset being used for training and 30% for testing. And the next step was Scale Invariant Feature Transform (SIFT) is used to extract shape features, colour moments are used to represent colour, and Segmentation-Based Fractal Texture Analysis (SFTA) is used to describe texture features. Because local demand continues, this paper's experiments are limited to a small variety of Malaysian medicinal herbal plants. And also, any publicly available image dataset, such as the Flavia Dataset, the SmithSonian Leaf Dataset, the Swedish Leaf Dataset, or the Image CLEF dataset, should be used to test the proposed methods. Other low-level feature extraction methods should also be tried to ensure consistency of results.

Hence, for performance evaluation, a basic identification rate is utilised. Before achieving definitive conclusions, additional evaluation measurements such as the recall-precision curve, F-measure, or accuracy should be performed. The main disadvantage of this model is that it is difficult to learn a number of multiple datasets.

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.

Color-based segmentation outperforms other published segmentation methods in part because the constraints for this task, which include preserving fine-scale features like thin structures and operating at interactive speeds, are more stringent than commonly assumed in the literature, but are common in many application domains. Curvature histograms have been demonstrated to be effective shape descriptors, in part because they can be generated reliably using integral measures. More collaborative capabilities, such as the ability to utilize the app with full classrooms,could be added to the software.

2.6 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. And, to our knowledge, Android is more popular than IOS as a mobile operating system. And also, this research describes an Android-based Smartphone application that uses photos of tree leaves to automatically identify plant species. One leaf image in this application can be a digital image from an existing leaf image database or a photograph taken with a camera. A single leaf should be put on a light, untutored background with no extra distractions. Leaf picture segmentation, feature extraction, and species identification are the three processes in the identification procedure.

The next was, the demo system is tested against the ImageCLEF2012 Plant Identification database, which contains 126 tree species native to the French Mediterranean region. The top several species that best match the query leaf image, as well as written descriptions and other photos of plant leaves, flowers, and so on, are the system's outputs to users. Our 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.

2.7 MULTI-ORGAN PLANT IDENTIFICATION

Israa Mohammed Hassoon et al. (2019) [2], proposed a review of various methods of plant species identification offered by. We concentrated on plant identification methods that relied on leaves, blooms , and fruits rather than seeds or branching style. The primary purpose of this paper is to describe various plant identification approaches, feature extraction methods, and how they are mathematically represented.

Plant recognition or classification can be done based on plant parts such as: 1) Plant Leaves 2) Plant Flowers 3) Plant's Fruits Plant Recognition Methodologies 1) Leaf-based approaches: While most studies extract data from leaf shape or colour, there are several essential leaf properties such as "aspect ratio," "narrow factor," "compactness," "centroid," "eccentricity," "dispersion," "area," "equivalent diameter," "moments invariant," and so on. 2) Methods based on flower or fruit shape characteristics such as "area," "perimeter," "roundness," and "aspect ratio". Techniques for Extraction and Representation of Features: a) There are two types of shape representation: "boundary-based" and "region-based." Texture (b) I Model-based method ii) Structural iii) Statistical Analysis: iv) Matrix of gray-level co-occurrence b) Color is more important in flower analysis than in leaf analysis. d)Venation in these vein features must be removed without affecting the contour of the leaf; while using a scanner, it was required to use a cleaned leaf image.

The review's main findings are as follows: Most studies based on leaf shape analysis, region-based description is less popular than boundary-based description, leaf analysis, shape and texture are considered more discriminative than colour, cleaned leaves with good quality were used in leaves database, most researches avoided segmentation, images were used with plain background, most plant identification approaches focus on leaves, shape is the preeminent feature for plant identification, most studies based on leaf shape analysis, region-based description is less popular than boundary-based description, region-based description is less popular than boundary. The main disadvantage is, this model takes only leaves for identify the species of plant.

Neha Goyal et al. (2018) [5], recommended various feature extraction approaches, classification, and other difficulties are used to identify plant species. Furthermore, the precision, recall, and F-Score of two widely used classifiers, Support Vector Machine (SVM) and Probabilistic Neural Network (PNN), are evaluated on the Flavia dataset. It has been discovered that SVM outperforms PNN when it comes to identifying plant species. Automatic plant species identification is a new topic of study with various important applications. A study of plant species identification techniques in depth. In terms of feature extraction methods, classifiers, and stated accuracy, various methods in the literature are compared. We may conclude from the review that the majority of studies are based on leaf shape and edge.

In certain research, the vein characteristic is substituted for the form. Furthermore, we have done tests based on various feature extraction approaches, classification, and other problems utilizing two well-known classifiers, namely PNN and SVM, for Automatic Plant species identification methods. Furthermore, the precision, recall, and F-Score of two widely used classifiers, Support Vector Machine (SVM) and Probabilistic Neural Network, are evaluated on the Flavia dataset. It has been discovered that SVM outperforms PNN when it comes to identifying plant species. Automatic plant species identification is a new topic of study with various important applications. A study of plant species identification techniques in depth. In terms of feature extraction methods, classifiers, and stated accuracy, various methods in the literature are compared. We may conclude from the review that the majority of studies are based on leaf shape and edge. In certain research, the vein characteristic is substituted for the form. The main disadvantage is this model consists of more complexity.

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.

This research also looked into a deep learning approach for discovering the representations needed for classification automatically, allowing for the usage of a single end-to-end pipeline for recognizing plants in the wild. In the test set, the suggested model ResNet26 achieves 91.78 percent accuracy, suggesting that deep learning is a promising method for large-scale plant categorization in natural settings. 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.

Jana Waldchen et al. (2017) [12], proposed the first comprehensive literature review with the goal of a complete analysis and comparison of primary studies on computer vision algorithms for plant species identification. We found 120 peer-reviewed research published in the last ten years (2005–2015) that were chosen through a multi-stage procedure. We present the applicable methods categorized according to the researched plant organ and the studied aspects, i.e. shape, texture, colour, margin, and vein structure, after a detailed review of these studies. We also assess approaches based on the accuracy of categorization performed on publicly available datasets. Our findings are significant to ecological and computer vision research, both of which are ongoing. A classification process can be separated into two parts: 1) picture acquisition and 2) image categorization. 2) Pre-processing 3) Extraction and description of features 4) The classification system. The goal of this paper was to locate, analyze, and compare research on plant species identification using computer vision algorithms.

A systematic review was carried out, guided by research questions and utilizing a well-defined data extraction and analysis process. The findings summarize the major findings of this systematic study and suggest research directions for the future. 1) The majority of investigations are carried out by computer scientists. 2) There are only two methodologies that have been tested on huge datasets. 3) The majority of research avoided segmentation by using photos with a plain background. 4) The main emphasis of study for plant identification is leaf analysis. 5) Shape is the most important feature for plant identification. 6) Multi-feature fusion allows for improved classification accuracy. 7) Shape descriptions based on contours are more common than those based on regions. 8) It is extremely difficult to compare and evaluate offered methods. The advantage of this model is it explains very well and is easy to understand and the disadvantage is this model takes more time.

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.

Javed Hossain et al, (2010) [23] suggested research provides a simple and computationally efficient leaf image-based technique for plant species recognition. Only plants with broad flat leaves that are more or less two-dimensional in nature operate with this strategy. The approach is divided into five sections. To begin, digital cameras or scanners are used to capture photos of the leaf. The user next chooses the leaf's base point as well as a few reference locations on the leaf blades. The leaf shape is extracted from the background using these points, and a binary image is created. The leaf is then horizontally aligned with its base point on the image's left.

The eccentricity, area, perimeter, main axis, minor axis, equivalent diameter, convex area, and extension of the object are then extracted.Slicing across the major axis and parallel to the minor axis extracts a unique set of features from the leaves. The feature points are then normalized by dividing the slice lengths by the leaf lengths (major axis). These characteristics are fed into the probabilistic neural network as inputs. 1200 basic leaves from 30 distinct plant species were used to train the network. Leaf identification, Plant biometrics, Plant recognition, and Probabilistic neural network are some of the terms used in this article.The advantage of the system has a recognition accuracy of 91.41% on average. Our system's major flaw is that it relies on user assistance during the pre-processing step.

Another drawback is that it is unable to cope with photographs that have a complex background. The user next chooses the leaf's base point as well as a few reference locations on the leaf blades. The leaf shape is extracted from the background using these points, and a binary image is created. The leaf is then horizontally aligned with its base point on the image's left. 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. We also assess approaches based on the accuracy of categorization performed on publicly available datasets. The advantage of this model is it explains very well and is easy to understand and the disadvantage is this model takes more time.

The user next chooses the leaf's base point as well as a few reference locations on the leaf blades. The leaf shape is extracted from the background using these points, and a binary image is created. The leaf is then horizontally aligned with its base point on the image's left. 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. We also assess approaches based on the accuracy of categorization performed on publicly available datasets. The advantage of this model is it explains very well and is easy to understand and the disadvantage is this model takes more time. The leaf shape is extracted from the background using these points, and a binary image is created. The leaf is then horizontally aligned with its base point on the image's left.

2.8 COMPARISON TABLE OF EXISTING ALGORITHM

S.NO	Author and Year	Methodology	Merits	Demerits
1	Amala Sabu et al, 2017	Identification of medicinal plants from their leaves which makes use of OpenCV and machine learning.	Features that are computationally difficult to be extracted are achieved.	The leaf dataset considered is small and instead of using a K-NN classifier, using SVM or ANN classifier gives a better result.
2	Sharifalillah Nordin et al, 2013	Scale-Invariant Feature Transform (SIFT) for shape feature extraction, Grid-Based color Moment (GBCM) for color feature extraction	The data covers 126 tree species and 11572 images which is a vast number as it scores a good accuracy.	Accuracy can be improved by adding a robust feature such as texture and also intend to experiment on learning classifiers such as supervised classifiers to increase the identification rate.
3	Neha Goyalet et al, 2018	Plant classification using SVM & PNN are compared in terms of precision, recall &f-score	Classification based on leaf shape, margin, and veins of leaves.	It does not work well with the larger dataset. Features of leaves are not included which is recognition of identification.
4	Parul Mittal et al, 2018	Plant classification based on physical attributes, SVM, Naive, and decision tree classifier	The accuracy is intensified as compared to existing algorithms.	The method doesn't include all physical attributes.
5	S. Anubha Pearline et al, 2018	Plant species recognition is carried out using two methods a) Traditional b) Deep Learning and classifiers like CNN-based pre-trained models.	Recognition using shape texture, color, and standard dataset provides good accuracy.	The accuracies can be improved by increasing the number of images using dataset augmentation methods.
6	EM Imah et al, 2018	Plant species identification using GRLVQ algorithm	Has higher accuracy than Random Forest and SVM has the fastest CPU time for plant identification.	A digital image has complex data if the quality varies the output decreases.
7	Aparajita Sahay et al, 2016	Plant species recognition using k-nearest neighbor	Recognition of plant from a simple photo through a product.	The classification of a leaf is based on the background we place a leaf. If the background (or) the surface color of the place where we place a leaf is not white then the results may differ.
8	Zhing-Qiu Ma et al, 2014	Leaf image segmentation feature extraction species extraction	Recognition of plants using photographs of plant leaves.	The dataset is specific to area the background threat is present at one time one image to be processed. The language used is java.
9	Esraa EWiharini et al, 2014	Linear Discrimination Analysis (LDA) algorithm	LDA has high accuracy which classifies using shape, color, texture, margin.	Data set is limited and if big data is to be tested then the testing would be complex.
10	Luciano DS Pacifico et al, 2018	Multi-layer perception artificial neural network (MPANN) trained with back propagation algorithm	It's easy to implement as the language used is python.	The dataset used consists of only iris, wheat seeds and 100 plant leaves of same plant i.e. dataset is limited.

Table 2.1 Comparision Table of Existing System

2.9 SUMMARY

After carefully surveyed all the reference papers of plant identification and classification, it is identified that there are still some accuracy issues present in those papers and none discussed plant classification based on seasons. So, to overcome this issue, this proposed method worked on plant classification and identification using tensorflow decision tree algorithm.

CHAPTER 3

SYSTEM REQUIREMENTS

3.1 PREAMBLE

The previous chapter analyse the many research articles and journals, their merits, demerits, and future directions, as well as the literature review connected to multimodal template protection reviews and various approaches and strategies of template protection. This chapter discusses the many tools and requirements that were utilised to construct the proposed system. This work includes information on the tools utilised, such as front-end tools, back-end tools, hardware tools, and other system requirements, such as hardware and software.

3.2 FRONT END TOOL

3.2.1 Jupiter Notebook

Jupiter is a free open-source interactive web tool known as a computational notebook that allows researchers to compile all of the software code, computational output, explanatory text, and multimedia resources into a single document or file. Jupiter Notebook is an easy-to-use interactive data science and environment that works with a variety of programming languages and may be used as an IDE, a presentation tool, or an educational tool. Jupiter notebook is not included with python, therefore if you wish to use it, you'll need to install Jupiter first, followed by any python distribution.

Figure 3.1: Jupiter Notebook

The Jupiter notebook is made up of two parts. In a front-end web page, users enter computer code or text into rectangular cells. After that, the browser sends the code to a back-end 'kernel,' which executes it and provides the results.

3.2.2 Anaconda IDE

Anaconda IDE is used for the implementation of entire project.

- **Anaconda**

Anaconda is a free and open-source Python and R programming language distribution for scientific computing, machine learning applications, large-scale data processing, predictive analytics, and more, with the goal of simplifying package management and deployment. Conda, the package management system, keeps track of package versions. Data-science packages for Windows, Linux, and macOS are included in the Anaconda installation. Anaconda Navigator is a desktop graphical user interface (GUI) that comes with the Anaconda distribution that allows users to run programmes and manage conda packages, environments, and channels without having to use command-line commands. Navigator may search for packages on Anaconda Cloud or in a local Anaconda Repository, then install, execute, and update them in an environment. Anaconda is an open-source package manager and environment management system that instals, runs, and updates packages and their dependencies across multiple platforms and languages.

3.3 COMMUNICATION INTERFACE

Python is a high-level programming language that is interpreted, interactive, and object-oriented. It was created between 1985 and 1990 with the cooperation of Guido van Rossum. Python source code is also available under the GNU General Public License, just like Perl (GPL). Python is a scripting language that is high-level, interpreted, interactive, and object-oriented. Python is written in a way that makes it extremely readable. It frequently uses English key phrases, whereas other languages utilise punctuation, and it has less syntactical structures than other languages**.**

- **Python is Interpreted :** Using the interpreter, Python is processed at runtime. You don't have to put your software together before running it. Perl and PHP are similar in this regard.

- **Python is Interactive :** You can sit down at a Python activate and begin coding right away with the interpreter.
- **Python is Object-Oriented :** Python supports the Object-Oriented programming approach, which encapsulates code within objects.
- **Python is a Beginner's Language :** Python is an excellent language for beginning
  programmers, as it allows for the creation of a wide range of programmes, from simple text processing to web browsers and games.

3.4 CHARACTERISTICS OF PYTHON

Python is a very simple language to learn compared to other programming languages such as c, c#, JavaScript, and java.

- The Python programming language is freely available on the official website and may be downloaded from www.python.org.
- It allows dynamic type checking and gives very high-level dynamic data records.
- It can be easily integrated into C, C++, COM, ActiveX, CORBA, and Java programmes.

3.5 HARDWARE REQUIREMENTS

- System: Intel or compatible Pentium class 4 processor or higher
- Hard Disk :1 TB
- Memory (RAM): At least 4GB
- OS: Windows 7/Windows 10
- System type: 64-bit Operating System
- Input devices: Keyboard, Mouse

3.6 SOFTWARE REQUIREMENTS

- Coding Language: Python
- IDE: Jupiter Notebook

3.7 ANACONDA INSTALLATION

- The following are the steps to install the Anaconda software Python3.6. They are

Step1: Download the Anaconda software click on Python3.6 to download

Step 2: Then click on agree button and then choose a location to download software and then click on the next button to continue.

Step 3: Click on Install and then click on Next button

Step 4: Successfully installation complete

Figure 3.2: Downloading Anaconda Software

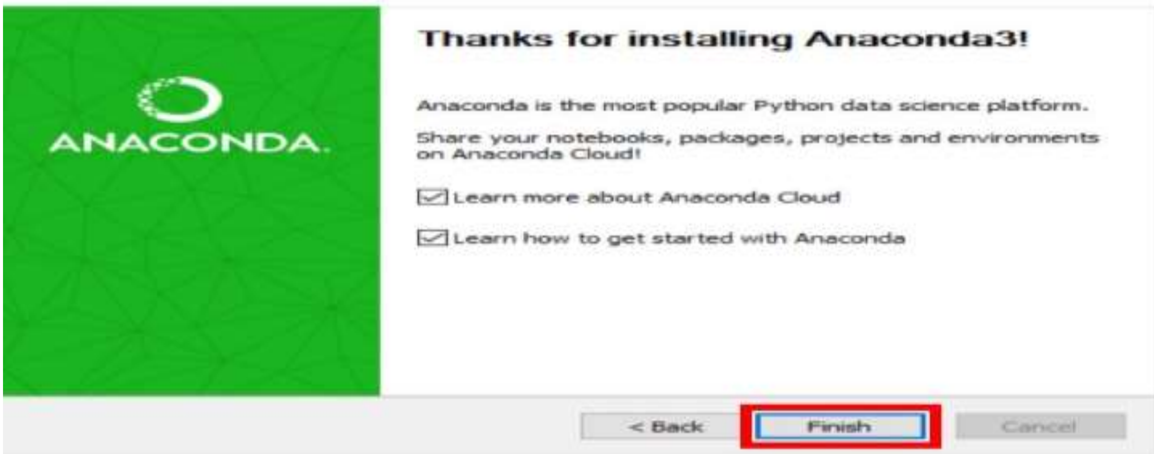

Figure 3.3: Installation Completed

3.8 PYTHON LIBRARIES

3.8.1 NumPy

NumPy is the most important Python package for scientific computing. NumPy arrays allow you to do advanced mathematical and other operations on massive amounts of data. Such actions are typically performed more quickly and with less code than utilising Python's built-in sequences**.**

3.8.2 Matplotlib

Matplotlib. Py plot is a set of utilities that allow matplotlib to behave similarly to MATLAB. Each py plot function modifies a figure in some way, such as creating a figure, a plotting area in a figure, charting certain lines in a plotting area, decorating the plot with labels, and so on.

3.8.3 Pandas

Pandas are primarily used to analyse data. Pandas supports importing data from a variety of file formats, including comma-separated values, JSON, SQL, and Microsoft Excel. Pandas supports a variety of data manipulation applications such as merging, reshaping, and selecting, as well as data cleaning and data wrangling features.

3.8.4 Scikit-Learn

Scikit-learn is a Python machine learning library that is free to use. It supports Python numerical and scientific libraries such as NumPy and SciPy, as well as various algorithms such as support vector machine, random forests, and k-neighbors.

3.8.5 Tensorflow

Tensor Flow is an open-source numerical computation library. Google invented and maintains it, and it's open-source under the Apache 2.0 licence. Although there is access to the underlying C++ API, the API is ostensibly for the Python programming language. TensorFlow was created with the intention of being used in both R&D and production systems. Google invented and maintains it, and it's open-source under the Apache 2.0 licence. Although there is access to the underlying the API is ostensibly for the Python programming language.

3.8.6 Keras

Keras is a Python-based high-level neural network API that can be used on top of TensorFlow. It makes prototyping simple and quick (through user friendliness, modularity, and extensibility). Convolutional and recurrent networks, as well as mixtures of the two, are supported. Runs without a hitch on both CPU and GPU**.**

3.8.7 Label Encoder

Label encoding is the process of converting labels into a numeric format so that they can be read by machines. Machine learning algorithms can then make better decisions about how those labels should be used. In supervised learning, it is an important pre-processing step for the structured dataset.

3.8.8 Train-test-split

When machine learning algorithms are used to make predictions on data that was not used to train the model, the train-test split procedure is used to estimate their performance.

- **Model.score :** Score(X train ,Y train) measures the model's accuracy against the training data. (How well the model explains the data with which it was trained.)
- **Model.predict :** Predict the label of a new set of data using a trained model. This method takes a single argument, the new data X new (for example, model. predict(X new)), and returns the learned label for each object in the array.

3.8.9 Decision tree classifier

Decision Trees are a type of Supervised Machine Learning (you explain what the input is and what the corresponding output is in the training data) in which the data is continuously split based on a specific parameter.

3.9 DATASET

The dataset considered in this model is our own customized dataset of 1340 species. It contains fruits, flowers, vegetables and herbs. The parameters considered in the dataset are name of the plant, season, minimum and maximum temperature, minimum and maximum humidity, and soil type.

3.10 SUMMARY

This chapter introduced the system's software and hardware requirements. The above requirements are required for the project's successful implementation. Python is the programming language of choice for the implementation of compact and secure multimodal template protection schemes that employ deep hashing for improved implementation and accuracy. Most Jupiter notebooks are in use. A system running the

64-bit Windows operating system and containing an Intel core processor device. The hardware requirements are the same as those required for running any regular Python script.

CHAPTER 4

SYSTEM DESIGN

4.1 PREAMBLE

The previous chapter describes the different requirements and tools used for developing the proposed system. It also consists of the hardware requirements like processor, memory, and software requirements like python and frontend and backend tools. This chapter deals with the design of the proposed system. It includes the system architecture, block diagram, and various diagrams to represent the design of the system.

4.2 PROPOSED SYSTEM BLOCK DIAGRAM

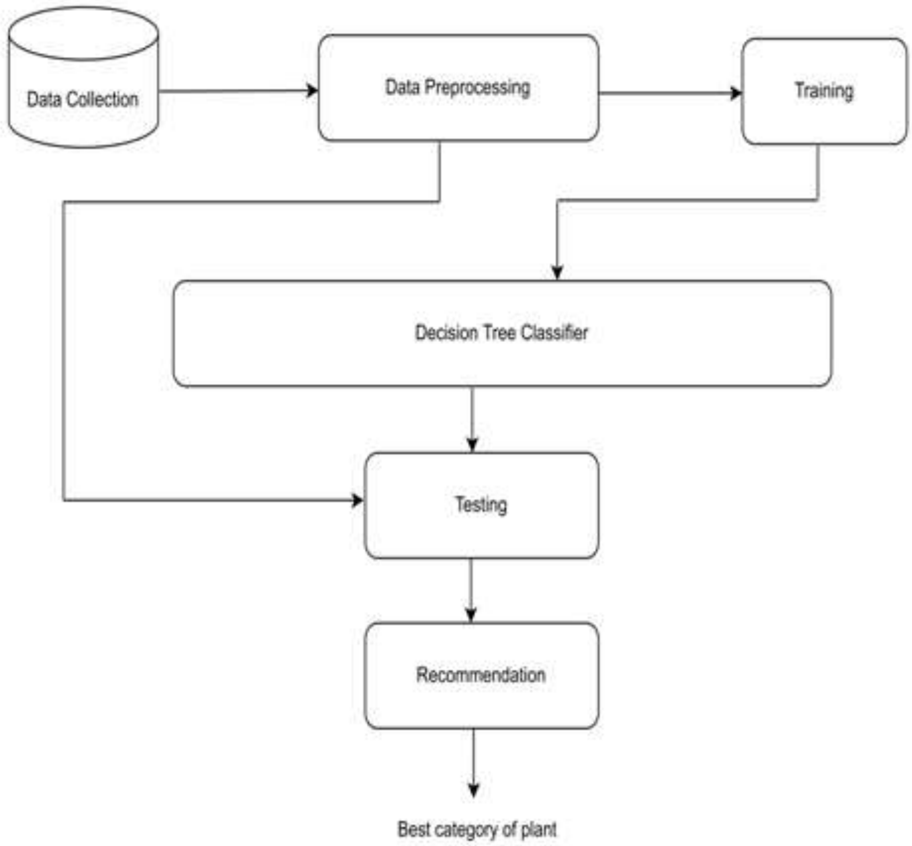

Figure 4.1: Flowchart of proposed system

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.

4.3 DATA PREPROCESSING

Data preprocessing 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 preprocessing is to ensure that the data is of good quality. The quality of the data can be checked by following factors.

- **Accuracy**

Pre-processing of data can sometimes result in unexpected improvement in the model accuracy. The degree to which data accurately portrays a real-world circumstance and is confirmed by a verifiable source is known as data accuracy. Data accuracy ensures that the real-world entities involved may participate as anticipated. It checks whether the data entered is correct or not.

- **Completeness**

Depending on the entity, this dimension can cover a wide variety of properties. The completeness of the data determines whether it is sufficient to make useful inferences and decisions.

- **Consistency**

This dimension indicates if the same data is recorded and used in multiple places. It's expressed as a percentage of values that match across multiple records. Data consistency guarantees that analytics collect and utilize the value of data accurately. Data consistency is generally linked to data accuracy, therefore any data set that scores well on both will be considered high-quality.

- **Validity**

This dimension denotes the availability of value characteristics for alignment with a certain domain or demand. Any invalid data will have an impact on the data's completeness. To ensure completeness, rules can be defined to ignore or resolve invalid data.

- **Uniqueness**

This dimension shows if the data set utilized has a single documented instance. The most important factor in avoiding duplication and overlap is uniqueness. The uniqueness of data is determined by comparing all records in a data collection or across data sets. A high uniqueness score ensures that duplicates and overlaps are minimized, resulting in data and analysis that can be trusted. Identifying overlaps can help keep records unique, while data cleansing and deduplication helps get rid of duplicates. Data uniqueness also helps with data governance and compliance.

- **Integrity**

Integrity indicates that the attributes are maintained correctly, even as data gets stored and used in diverse systems. Data integrity ensures that all enterprise data can be traced and connected. Data transformation across systems can affect its attribute relationships.

- **Timeliness**

The value of data lies in its application. If data is not provided in a timely manner, it is worthless. The term timeliness refers to whether or not data is available when it is needed.

4.4 STAGES OF DATA PREPROCESSING

Data preprocessing is the procedure for preparing raw data for use in a machine learning model. It's the first and most important stage in building a machine learning model. Real-world data usually contains noise, missing values, and is in an unsuitable format that cannot be used directly in machine learning models. Data preprocessing is a necessary task for cleaning data and making it suitable for a machine learning model, which improves the model's accuracy and efficiency. Data preprocessing is divided into four stages: data cleaning, data integration, data reduction and data transformation.

The below image shows the stages of data preprocessing. In each stage there will be few issues and these issues were solved by some approaches.

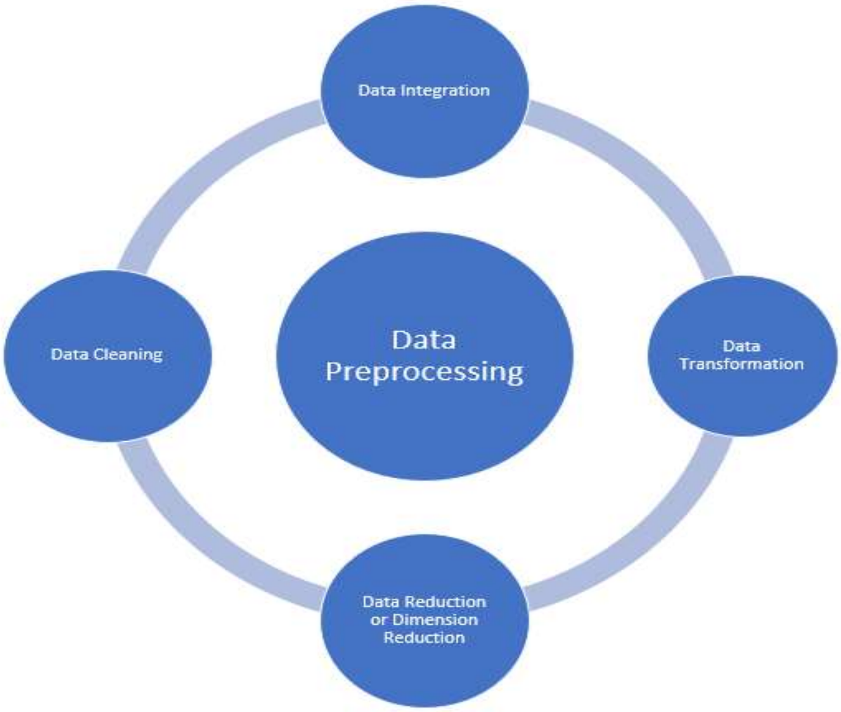

Figure 4.2: Stages of Data Preprocessing (analyticsvidhya.com)

4.4.1 Data cleaning

Data cleaning refers to techniques to 'clean' data by removing outliers, replacing missing values, smoothing noisy data, and correcting inconsistent data. Each of these activities is carried out using a variety of ways, each of which is tailored to the user's preferences or problem set.

Multiple approaches are there to deal with missing values. Removing the training example, filling in missing value manually, using a standard value to replace the missing value, using central tendency like mean, median and mode for attribute to replace the missing value, using central tendency measures for attribute belonging to same class to replace the missing value, using the most probable value to fill in the missing value these approaches are used to deal with multiple data. Noise is defined as a random variance in a measured variable. Each of these activities is carried out using a variety of ways, each of which is tailored to the user's preferences or problem set. To deal with these anomalous values, data smoothing techniques are applied like binning, regression and outlier analysis.

4.4.2 Data integration

Data integration has become a crucial aspect of the process because data is acquired from numerous sources. This may result in duplicate and inconsistent data, resulting in poor data model accuracy and speed. Approaches like tuple duplication detection and data conflict detection are being looked after to cope with these difficulties and maintain data integrity. The most frequent data integration approaches are data consolidation, data propagation and data virtualization.

4.4.3 Data reduction

The goal of data reduction is to provide a condensed version of the data set that is less in size while keeping the original data set's integrity. As a result, efficient but equivalent results are obtained. A few methods to reduce the volume of data are Missing values ratio, Low variance filter, High correlation filter and Principal component analysis. PCA method only works for features with numerical values.

4.4.4 Data transformation

The final step of data pre-processing is transforming the data into a form appropriate for data modelling. The following are some data transformation strategies: smoothing, attribute or feature construction, aggregation, normalization, discretization and Concept hierarchy generation for nominal data.

4.5 DIFFERENT TYPES OF RECOMMENDATION BASED ON INPUTS

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.

4.6 SUMMARY

The design of the system is explained in chapter 4. This chapter provides 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 next chapter describes the implementation details like modules or steps in implementation and proposed approach, techniques, and equations.

CHAPTER 5

IMPLEMENTATION

5.1 PREAMBLE

The previous chapter describes the design of the proposed system. It deals with the design phase consisting of system architecture. The next step after designing is the methodology of the proposed approach. This chapter deals with the methodology of the proposed approach. This chapter consists of the proposed approach and various steps required to implement the proposed approach. This chapter gives a brief view of the methodology of the system using various steps and proposed approaches**.**

5.2 PROPOSED METHODOLOGY

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.

5.3 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:

1. Dataset collection
2. Dataset Pre-processing
3. Splitting the dataset
4. Applying Machine learning algorithms

5.4 DATASET COLLECTION

The data collection process involves the selection of quality data for analysis. The dataset considered in this model is our own customized dataset. It contains fruits, flowers, vegetables, herbs and common dataset containing all categories of plant. The parameters considered in the dataset are name of the plant, season, minimum and maximum temperature, minimum and maximum humidity, and soil type. The below is the sample dataset

1	Name	Season	MinTemp	MaxTemp	MinHumidity	MaxHumidity	Soil
2	Blackpepper	rainy	23	32	50	60	red
3	Cinnamon	monsoon	20	30	60	70	brown
4	Alovera	monsoon	18	30	10	40	black
5	Ginger	summer	21	25	60	70	black
6	Sandalwood	monsoon	12	35	50	60	red
7	Ashwagandha	winter	20	35	90	100	red
8	Red clover	summer	20	30	60	70	black
9	Burdock	winter	20	25	75	80	black
10	Bayberry	winter	25	28	85	90	black

Table5.1 Sample dataset

The merged dataset contains all the four dataset records i.e. fruits, flowers, herbs and vegetables.

5.5 DATASET PRE-PROCESSING

The purpose of pre-processing is to convert raw data into a form that fits machine learning. Structured and clean data allows a data scientist to get more precise results from an applied machine learning model.

The technique includes encoding the dataset values.The dataset consist of 16 diffferent colors of soils and the category of plants we work in are Flower,Fruit,Herb,Vegetable.The soils are black,blue,brown,dark black,light black,light brown,light grey,light red,light yellow,loamy,red,reddish,reddish brown,sandy,yellow.

Encoded Value for Soil	Color 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 5.2 Encoded values for soil type

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

Table 5.3 Encoded values for plant category

Encoded Value for Seasons	Name of the Seasons
0	autumn
1	common
2	fall
3	monsoon
4	rainy
5	spring
6	summer
7	winter

Table 5.4 Encoded values for season

5.6 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 modelTest 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.

5.7 APPLYING MACHINE LEARNING ALGORITHM

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.

In this project Decision Tree algorithm is used for prediction. In which the model predicts which category of plant is best suitable to grow for the given input conditions.

5.7.1 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.

5.7.2 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.

5.7.3 Decision Tree Representation

Instances are classified using decision trees by sorting them down the tree from the root to a leaf node, which provides the classification. As indicated in the above diagram, an instance is classified by starting at the root node of the tree, checking the attribute specified by this node, and then progressing along the tree branch according to the attribute value. The subtree rooted at the new node is then processed in the same way.

The advantages of decision tree methods are decision trees can provide rules that are easy to understand. Classification is performed using decision trees, which do not require a lot of computing. Both continuous and categorical variables can be handled by decision tree. Decision trees show which fields are most essential for classification or prediction.

Terminology of decision tree is given as a vector of features or qualities that define the input space is referred to as instances. A quantity that describes an instance is called an attribute. The function that converts input to output. The function we're looking for, i.e. the real solution, is known as the target concept. Hypothesis Class: This is a collection of all conceivable functions. A sample is a sequence of inputs accompanied by a label indicating the proper output (also known as the Training Set) Notion of the Candidate: A concept that we believe is the target concept. Testing Set: A testing set is similar to a training set and is used to evaluate a candidate's concept and performance.

5.8 PARAMETERS

The Parameters used in this project are Temperature, Humidity, Season and Soil Type.

5.8.1 Temparature

Temperature is a physical number that describes how hot or cold something is. When a body comes into contact with another that is colder or hotter, it is the manifestation of thermal energy, which is present in all matter and is the source of the occurrence of heat, a flow of energy. Heat should not be confused with temperature.

A thermometer is used to determine the temperature. Thermometers are calibrated in a variety of temperature scales that have traditionally defined temperature using a variety of reference points and thermometric substances. The Celsius scale (previously known as centigrade, designated as °C), the Fahrenheit scale (denoted as °F), and the Kelvin scale (denoted as K) are the most prevalent scales. The Kelvin scale (denoted as K) is primarily used for scientific purposes under conventions of the International System of Units (SI).

Absolute zero is the lowest theoretical temperature at which no additional thermal energy can be collected from a body. The third law of thermodynamics recognises that it can only be approached extremely precisely (100 pK) but not reached experimentally.

Temperature is significant in physics, chemistry, Earth science, astronomy, medicine, biology, ecology, material science, metallurgy, mechanical engineering, and geography, as well as most elements of everyday life.

In this work temperature is divided as minimum temperature and maximum temperature. Thermometers are calibrated in a variety of temperature scales that have traditionally defined temperature using a variety of reference points and thermometric substances. Thermometers are calibrated in a variety of temperature scales that have traditionally defined temperature using a variety of reference points and thermometric substances. Thermometers are calibrated in a variety of temperature scales that have traditionally defined temperature using a variety of reference points and thermometric substances

5.8.2 HUMIDITY

Humidity refers to the amount of water vapour in the air. The gaseous condition of water, known as water vapour, is often invisible to the naked eye. [1] The presence of precipitation, dew, or fog is indicated by humidity.

Humidity is determined by the system's temperature and pressure. Cool air has a higher relative humidity than warm air when the same amount of water vapour is present. The dew point is a related metric. As the temperature rises, the amount of water vapour required to reach saturation rises as well. When the temperature of a parcel of air drops below a certain level, it will eventually approach saturation without adding or losing water mass. 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 (in grammes per cubic metre)[2] or the mass of water vapour per mass of dry air are used to calculate absolute humidity (usually in grammes per kilogram). [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 misleading, 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. There is a little discrepancy indicated below under "Enhancement factor," which can be overlooked in many computations unless extreme precision is desired. As the temperature rises, the amount of water vapour required to reach saturation rises as well When the temperature of a parcel of air drops below a certain level, it will eventually approach saturation without adding or losing water mass. The idea of air holding or being saturated by water vapour is frequently emphasised in relation to the concept of relative humidity.

5.8.3 SEASON

A season is a division of the year depending on weather, ecology, and the number of daylight hours in a particular place. Seasons on Earth are caused by the Earth's orbit around the Sun and its axial tilt in relation to the ecliptic plane. Seasons in temperate and polar regions are defined by differences in the amount of sunshine that reaches the Earth's surface, which can lead animals to hibernate or migrate, and vegetation to become dormant. Various civilizations define the number and type of seasons based on regional differences, and as a result, the number of seasons differs among present and historical cultures**.**

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

Spring, summer, autumn (or fall), and winter are the four seasons acknowledged by the Gregorian calendar in temperate and sub-polar countries. For temperate climate regions with no fixed calendar dates, ecologists frequently adopt a six-season model: prevernal, vernal, estival, serotinal, autumnal, and hibernal. There are two seasons in many tropical regions: the rainy, wet, or monsoon season, and the dry season. Some people have a third season that is cool, mild, or harmattan.

The time of major biological phenomena such as hurricane season, tornado season, and wildfire season can also determine "seasons." [requires citation] The ancient Egyptian seasons of flood, growth, and low water, which were historically defined by the annual flooding of the Nile in Egypt, are examples of historical significance. Seasons on Earth are caused by the Earth's orbit around the Sun and its axial tilt in relation to the ecliptic plane. Seasons in temperate and polar regions are defined by differences in the amount of sunshine that reaches the Earth's surface, which can lead animals to hibernate or migrate, and vegetation to become dormant.There are two seasons in many tropical regions: the rainy, wet, or monsoon season. and the dry season. Seasons in temperate and polar regions are defined by differences in the amount of sunshine that reaches the Earth's surface, which can lead animals to hibernate or migrate, and vegetation to become dormant.

5.8.4 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 sometimes known as earth or dirt; some scientific definitions differentiate dirt from soil by limiting the former term to displaced soil alone.

The lithosphere, hydrosphere, atmosphere, and biosphere all interact with the pedosphere.

The term pedolith, which is usually used to refer to soil, means "ground stone" in the sense of "basic stone," and comes from the ancient Greek v, which means "ground, earth." Soil is made up of two phases: a solid phase made up of minerals and organic matter (the soil matrix) and a porous phase that retains 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.

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. It develops throughout time as a result of a variety of physical, chemical, and biological processes, including weathering and erosion.

5.9 ALL SEASONS

The vector of features or qualities that define the input space is referred to as instances. A quantity that describes an instance is called an attribute.

The concept is the function. Seasons serve as a reminder that change is both a natural law and a sign of development. According to the ancient Hindu calendar, India has six seasons (the Lunisolar Hindu). The twelve months of the year are divided into six seasons, each lasting two months.

- **Vasant Ritu (Spring)** : The spring season in India is traditionally defined as the months of March and April, when the average temperature is around 32 degrees Celsius. Some people in India, particularly in the state of Karnataka, celebrate Ugadi, the Hindu New Year, in the spring.
- **Grishma Ritu (Summer):** Temperature - Summers in Andhra Pradesh are hot. On the hottest days, temperatures can reach 45°C, with daytime temperatures averaging 40°C. The temperature drops in the evening, but it remains around 28°C. During the summer, Andhra Pradesh is dominated by the heat.Minimum/Maximum Temperature July - September: 25 - 32°C.Temperatures range from 28 to 45 degrees Celsius from March through June. Temperatures range from 16 to 25 degrees Celsius from October to February.
- **Varsha Ritu (Monsoon):** The monsoon brings relief. Temperatures are approximatel 35 degrees Celsius (95 degrees Fahrenheit), although humidity is considerable; nights are around 27 degrees Celsius (81 degrees Fahrenheit). This is when the majority of the rain falls, which can result in serious flooding. During the rainy season, the sun is frequently obscured.
- **Sharad Ritu (Autumn):** Temperatures in the fall range from 72.7 degrees Fahrenheit (22.6 degrees Celsius) in Florida to 26.7 degrees Fahrenheit (-2.9 degrees Celsius) in Alaska. The season averages 53.9 °F (12.2 °C) across the United States, excluding Hawaii and Alaska.
- **Hemant Ritu (Pre-Winter)** : Pre-winter (mid-October to December), also known as Hemant Ritu, is the period preceding winter. Pre-winter is the season that follows fall and before winter.

- **Shishir Ritu (Winter):** The minimum temperatures observed over the last three days have ranged from 20 to 24 degrees Celsius. In Andhra Pradesh, maximum temperatures will be below normal during the winter season. This is when the majority of the rain falls, which can result in serious flooding. During the rainy season, the sun is frequently obscured.Majority of the plants tends to be growing well in this particular season. Inspite of suitable conditions,having too low temperatures constantly plants growth improperly.

5.10 SUMMARY

In this chapter, we briefly discussed about the methodology and the algorithms that are used in this proposed work. Methodology is the main framework in the project and it is the main thing which gives the result of our work. So, using the best algorithms gives the best results and this is why we compare different algorithms which it leads to the success of the project.

CHAPTER 6

EXPERIMENTAL RESULTS

6.1 PREAMBLE

In the previous chapter, the proposed system's methodology for developing a Plant Recommendation Model for Urban People is discussed. This chapter discusses the experimental results of the proposed approach.

6.2 RESULTS

The 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.

```
In [12]: def pred():
             k=model.predict([[MinTemp,MaxTemp,MinHumid,MaxHumid,s,st]])
             k1=[0,1,2,3]
             k2=["Flower","Fruit","Herb","Vegetable"]
             if k in k1:
                 print(k2[k1.index(k)])
             else:
                 print("Enter Valid Inputs")
         if Season=="summer" and MinTemp<=15 and MaxTemp>=60:
             print("enter valid inputs")
         elif Season=="winter"and MinTemp<=-25 and MaxTemp>=30:
             print("enter valid inputs")
         else:
             pred()

         Herb
```

Figure 6.1: Prediction of plant category

```
Do you wish to continue: yes/no
yes
Type of plant:fruit
1:Recomendation using name and output is required conditions for that plant
2:Recomendation using season and output is ideal plants for that season
3:classification using name,temperature,humidity,soil and output is either yes or no
4:Recomendation using Temperature and Humidity,name and soil are outputs
choose one of the options (1/2/3/4): 1
Enter Name of the plant:mango
```

	Name	Minimum Temperature	Maximum Temperature	Minimum Humidity	Maximum Humidity	Season	Soil
0	Mango	23	30	50	95	summer	black

Figure 6.2: Recommendation using name and output is required conditions for that plant

The first recommendation is, if the name of the plant is given as input then the required conditions that are needed to grow the plant are given as output. The recommendation phase has four categories of outputs based on the inputs. A sample example is considered for the first recommendation model. The first recommendation is, if the name of the plant is given as input then the required conditions that are needed to grow the plant are given as output.

The second recommendation is that if the season is given as input then the plants that are grown in that particular season are given as output.

```
Do you wish to continue: yes/no
yes
Type of plant:fruit
1:Recomendation using name and output is required conditions for that plant
2:Recomendation using season and output is ideal plants for that season
3:classification using name,temperature,humidity,soil and output is either yes or no
4:Recomendation using Temperature and Humidity,name and soil are outputs
choose one of the options (1/2/3/4): 2
Enter Name of the season: summer
Choose one of the following plants to opt in:

'Kashmiri Apple'

'Golden Apple'

'Thompson grapes'

'Dilkhush'
```

Figure 6.3: Recommendation using season and output is ideal plants for that season.

The classification is made by giving all parameters as input such as name of the plant, temperature, humidity and soil type then the model gives yes\no. Here, yes indicates plant can be grown and no indicates plant cannot be grown. The classification is made by giving all parameters as input such as name of the plant, temperature, humidity and soil type then the model gives yes\no. Here, yes indicates plant can be grown and no indicates plant cannot be grown.

```
Do you wish to continue: yes/no
yes
Type of plant:fruit
1:Recomendation using name and output is required conditions for that plant
2:Recomendation using season and output is ideal plants for that season
3:classification using name,temperature,humidity,soil and output is either yes or no
4:Recomendation using Temperature and Humidity,name and soil are outputs
choose one of the options (1/2/3/4): 3
Enter the Plant Name: mango
Enter the Average Temperature of your locality: 33
Enter the Average Humidity of your locality: 75
Enter the soil type do you check in: black.
Mango
Yes you can this Mango.The temperature 33^c and humidity 75% are in the range.
```

Figure 6.4: Classification using name, temperature, humidity, soil and output is either yes/no

The classification is made by giving all parameters as input such as name of the plant, temperature, humidity and soil type then the model gives yes\no. Here, yes indicates plant can be grown and no indicates plant cannot be grown.

The last recommendation is made by giving the temperature and humidity of the area as the inputs then the name of the plant and soil type which are suitable to grow in those conditions are given as outputs.

```
Type of plant:fruit
1:Recomendation using name and output is required conditions for that plant
2:Recomendation using season and output is ideal plants for that season
3:classification using name,temperature,humidity,soil and output is either yes or no
4:Recomendation using Temperature and Humidity,name and soil are outputs
choose one of the options (1/2/3/4): 4
Enter the temperature of your locality: 25
Enter the humidity of your locality: 70
plant :Mango                    soil :black
plant :Grapes                   soil :black
plant :Banana                   soil :brown
plant :Watermelon                       soil :dark brown
plant :Strawberry                       soil :black
plant :Sapota                   soil :brown
plant :Mosambi                  soil :brown
plant :Pear                     soil :brown
plant :Mangosteen                       soil :sandy
plant :Dragon fruit                     soil :brown
```

Figure 6.5: Recommendation using temperature and humidity, name and soil are outputs

This recommendation model is tested with different classifiers such as Decision Tree, Random Forest, K-NN and Support Vector Machine classifiers.

6.2.1 Decision Tree Classifier

It is a tree-structured classifier where internal nodes represent dataset attributes, branches represent decision rules and each leaf node provides the conclusion After training the dataset and applying Decision Tree classifier the classification accuracy of 97% is obtained.

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

Out[19]:  0.9694656488549618
```

Figure 6.6: Accuracy of Decision Tree Classifier

6.2.2 Support Vector Machine classifier

A support vector machine (SVM) is a machine learning algorithm for classification and regression analysis that examines data. SVM is a supervised learning approach that examines data and divides it into two groups. The output of an SVM is a map of the sorted data with the margins between the two as wide as possible. It is trained using a set of data that has already been divided into two categories, and it builds the model as it is trained. An SVM algorithm's job is to figure out which category a new data point fits in. As a result, SVM is classified as a non-binary linear classifier. After training the dataset and applying Support Vector Machine classifier the classification accuracy of 53% is obtained.

```
from sklearn.metrics import classification_report
print(classification_report(b_test,y_pred3))
```

```
              precision    recall  f1-score   support

           0       0.37      0.76      0.50        41
           1       0.53      0.47      0.50        64
           2       0.00      0.00      0.00        44
           3       0.65      0.70      0.67       113

    accuracy                           0.53       262
   macro avg       0.39      0.48      0.42       262
weighted avg       0.47      0.53      0.49       262
```

Figure 6.7: Accuracy of SVM Classifier

The Confusion Matrix for the Support Vector Machine Classifier is obtained as below.

```
In [15]: pd.crosstab(b_test,y_pred3)
Out[15]:
```

col_0	0	1	2	3
Name_n				
0	35	2	0	7
1	15	9	0	56
2	22	3	1	17
3	13	2	0	80

Figure 6.8: SVM Confusion Matrix

6.2.3 Random Forest Classifier

Random Forest is a classifier that combines a number of decision trees on different subsets of a dataset and averages the results to increase the dataset's predicted accuracy. Instead than relying on a single decision tree, the random forest collects the forecasts from each tree and predicts the final output based on the majority votes of predictions. The greater the number of trees in the forest, the more accurate it is and the problem of overfitting is avoided. After training the dataset and applying Random Forest classifier the classification accuracy of 95% is obtained.

```
In [64]:  from sklearn.metrics import classification_report
          print(classification_report(y_test,y_pred1))
```

	precision	recall	f1-score	support
0	1.00	0.89	0.94	55
1	0.94	0.97	0.95	63
2	0.95	1.00	0.98	40
3	0.94	0.96	0.95	104
accuracy			0.95	262
macro avg	0.96	0.96	0.96	262
weighted avg	0.96	0.95	0.95	262

Figure 6.9: Accuracy of Random Forest Classifier

6.2.4 K-Nearest Neighbor Classifier

The K-NN method assumes that the new case/data and existing cases are similar and places the new case in the category that is most similar to the existing categories. During the training phase, the KNN algorithm simply stores the dataset, and when it receives new data, it classifies it into a category that is quite similar to the new data. After training the dataset and applying K-NN classifier the classification accuracy of 82% is obtained.

	precision	recall	f1-score	support
0	0.80	0.78	0.79	45
1	0.89	0.79	0.84	78
2	0.66	0.98	0.78	41
3	0.89	0.79	0.83	98
accuracy			0.82	262
macro avg	0.81	0.83	0.81	262
weighted avg	0.83	0.82	0.82	262

Figure 6.10: Accuracy of K-nn Classifier

6.3 PERFORMANCE COMPARISON

The performance comparison for the graph is shown below

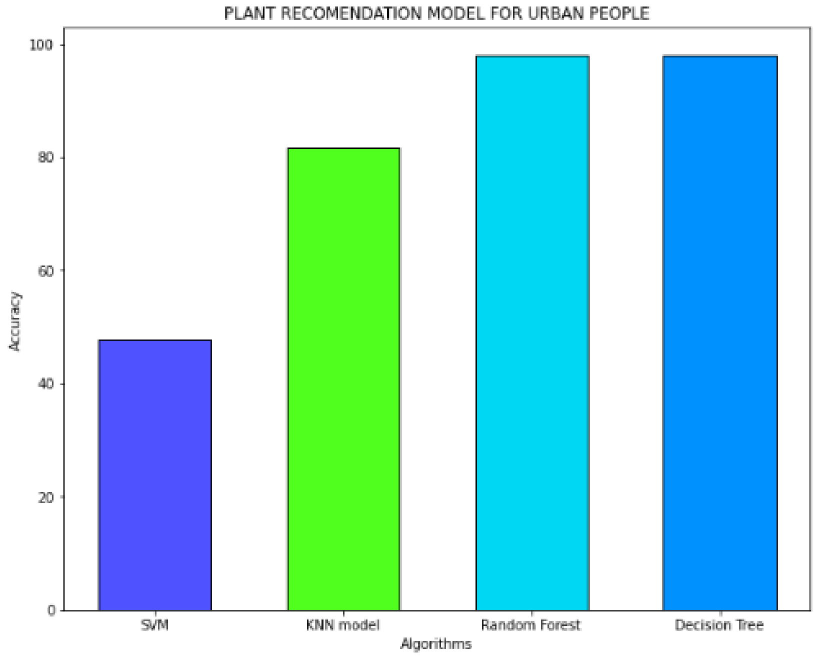

Figure 6.11: Performance Comparison Graph

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

Table 6.1 Performance Comparision Table

6.4 SUMMARY

This chapter discusses the experimental results of the proposed work and also explains the graphs and comparison results of the proposed system.

CHAPTER 7

CONCLUSION AND FUTURE ENHANCEMENT

7.1 CONCLUSION

This project "Plant Recommendation Model for Urban People" 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.

8.2 FUTURE ENHANCEMENT

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. 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.
15. Nursuriati Jamil, NurilAslina Che Hussin, SharifalillahNordin, Khalil Awang, "Automatic Plant Identification: Is Shape the Key Feature?", Procedia Computer Science, 2015.
16. Zhong-Qiu Zhao, Lin-Hai Ma, Yiu-mingCheung, Xindong Wu, Yuanyan Tang, Chun Lung Philip Chen, "ApLeaf: An efficient android-based plant leaf identification system", Elsevier, 2014.
17. EsraaElhariri, Nashwa El-Bendary, Aboul Ella Hassanien, "Plant Classification System based on Leaf Features", IEEE, 2014.
18. W.H. Rankothge, D.M.S.B Dissanayake, U.V.K.T Gunathilaka, S.A.C.M Gunarathna, C.M Mudalige, R.P Thilakumara, "Plant Recognition System Based on Neural Networks", ICATE, 2013.
19. Mohamad Faizal Ab Jabal, Suhardi Hamid, SalehuddinShuib, Illiasaak Ahmad, "Leaf feature extraction and recognition approaches to classify plant", Journal of Computer Science, 2013.
20. NurilAslina Che Hussin, Nursuriati Jamil, SharifalillahNordin, Khalil Awang," Plant Species Identification by using Invariant Feature Transform (SIFT) and Grid Based Colour Moment (GBCM)", IEEE, 2013.
21. A.H. Kulkarni, Dr.H.M. Rai, Dr.K.A. Jahagirdar, P.S. Upparamani, "A Leaf Recognition Technique for Plant Classification Using RBPNN and Zernike Moments", International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), 2013.
22. Neeraj Kumar, Peter N. Belhumeur, Arijit Biseas, David W. Jacobs, W. John Kress, Ida C. Lopez, Joao V.B. Soares, "Leafsnap: A Computer Vision System for automatic Plant Species Identification", Springer, 2012.
23. Javed Hossain, M. Ashraful Amin, "Leaf - Shape Identification Based Plant Biometrics", IEEE, 2010.
24. Thibaut Beghin , James S. Cope, Paolo Remagnino, Sarah Barman, "Shape and Texture Based Plant Leaf Classification", Springer, 2010.

25. Xiao-Feng Wang, De-Shuang Huang, Ji-Xiang Du, Huan Xu, Laurent Heutte, “Classification of plant leaf images with complicated background”, Applied Mathematics and Computation, 2008.
26. Stephen Gang Wu, Forrest Sheng Bao, Eric You Xu, Yu-Xuan Wang, Yi-Fan Chang, Qiao-Liang Xiang, “A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network”, IEEE, 2007.
27. Ji-Xiang Du, Xiao-Feng Wang, Guo-Jun Zhang, “Leaf shape- based plant species recognition”, Applied Mathematics and Computation, 2007.
28. JanezDemsar, “Statistical Comparisons of Classifiers over Multiple Data Sets”, Journal of Machine Learning Research, 2006.
29. Xiao Gu, Ji-Xiang Du, Xiao-Feng Wang, “Leaf Recognition Based on the Combination of Wavelet Transform and Gaussian Interpolation”, Springer, 2005.
30. Yanhua Ye, Chun Chen, Chun-Tak Li, Hong Fu, Zheru Chi, “A Computerized Plant Species Recognition System”, International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004.

APPENDIX A

SOURCE CODE

```
#Decision Tree Model

import pandas as pd

from sklearn.preprocessing import LabelEncoder

from sklearn import tree

model=tree.DecisionTreeClassifier()

df=pd.read_csv(r"D:\final_year_projects\imran\MergedDataset.csv")

# df=pd.read_csv(r"C:\Users\Imran\OneDrive\Desktop\nxtwave\mergeds.csv")

print(df)

le_Season=LabelEncoder()

le_Soil=LabelEncoder()

le_Name=LabelEncoder()

df["Season_n"]=le_Season.fit_transform(df["Season"])
df["Name_n"]=le_Name.fit_transform(df["Name"])
df["Soil_n"]=le_Soil.fit_transform(df["Soil"])
inputs=df.drop(["Name_n","Name","Season","Soil"],axis="columns")

target=df["Name_n"]

print(inputs)

print(target)

from sklearn.tree import DecisionTreeClassifier
model=DecisionTreeClassifier(criterion="entropy",random_state=0)
```

```
from sklearn.model_selection import train_test_split
inputs_train,inputs_test,target_train,target_test=train_test_split(inputs,target,test_size
= model.fit(inputs_train,target_train)

 #

inputs_test

MinTemp=int(input("Enter minimum temperature :"))

MaxTemp=int(input("Enter maximum temperature :"))

MinHumid=int(input("Enter minimum humidity :"))

MaxHumid=int(input("Enter maximum humidity :"))

Season=input("Enter the season:")

Season=Season.lower() s=0 s1=[0,1,2,3,4,5,6,7]
s2=["autumn","common","fall","monsoon","rainy","spring","summer","winter"]
dict_seas=dict(zip(s1,s2))

for k,v in dict_seas.items():

if v==Season:

s=k

else:

s=8

Soil=input("Enter the soil:")

Soil=Soil.lower()

st=0 l1=[0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15]

l2=["black","blue","brown","dark black","dark brown","light black","light
brown","light gre ,"red","reddish","reddish brown","sandy","yellow"]
```

```
if Soil in l2:

st=l1.index(l2.index(Soil))

print(st)

else:

st=16

k=model.predict([[MinTemp,MaxTemp,MinHumid,MaxHumid,s,st]])

k1=[0,1,2,3]

k2=["Flower","Fruit","Herb","Vegetable"]

if k in k1:

print(k2[k1.index(k)])

else:

print("Enter Valid Inputs")

print(k)

dt_train_ac=model.score(inputs_train,target_train)
dt_test_ac=model.score(inputs_test,target_test)

dt_ac=model.score(inputs_train,target_train)

print("Testing Accuracy is {}".format(dt_test_ac))

print("Training Accuracy is {}".format(dt_train_ac))

# from sklearn.metrics import classification_report, confusion_matrix

# print(confusion_matrix(target_train,target_test))

# print(classification_report(target_train,target_test))

import pandas as pd
```

```
target_predict=model.predict(inputs_test)

pd.crosstab(target_test,target_predict)

from sklearn.metrics import confusion_matrix,accuracy_score
y_predict=model.predict(inputs_test)

dt_accuracy=accuracy_score(target_test,y_predict)

print(dt_accuracy)

from sklearn.metrics import classification_report
print(classification_report(target_test,y_predict))

while(1):

print("Do you wish to continue: yes/no")

y=input()

y=y.lower()

if y=="yes":

import pandas as pd

#Recomendation using name of the plant and displaying required conditions to grow

def recomendation_using_name(name,dataset):

name=name[0].upper()+name[1:].lower()

for i in range(len(dataset)):

if name==dataset.Name[i]:

print("name: {}\n season :{}\n Minimum Temperature :{}^c\n Maximum Temp
.format(name,dataset.Season[i],dataset.MinTemp[i],dataset.MaxTemp[i],dataset

break
```

```
else:

print("Try another plant 😊👍")

#Recomendation using season and displaying all the possible growing plants in that season

def plant_recomendation_using_S(season,dataset):

lis=[] print("Choose one of the following plants to opt in:")

for i in range(len(dataset)):

if season==dataset.Season[i]:

lis.append(dataset.Name[i])

if len(lis)==0: print("Invalid entry")

else:

for i in lis:

print(i)

#Classification Function using label,temperature,humidity,soil parameters and concl

def plant_classification_using_LTHS(Label,Temperature,Humidity,soil,dataset):

Res=[] print(Label)

for i in range(len(dataset)):

if(Label in dataset.Name[i]):

Res.append("Y")

if Temperature>=int(dataset.MinTemp[i]) and
Temperature<=int(dataset.MaxTemp[i])

Res.append("Y")
```

```
if Humidity>=int(dataset.Minhumid[i]) and Humidity<=int(dataset.MaxHumid[i])
Res.append("Y")

if(soil==dataset.Soil[i]):

Res.append("Y")

#if any of the above cases fail then it will append "N"

else:

Res.append("N")

#conditional based classification

if(Res.count("Y")>=4):

print("Yes you can this {}.The temperature {}^c and humidity {}% are in the

else:

print("No choose some other plant,the conditions you inputs are not the best")

#Recomendation using Temperature and Humidity which is area specific plant
recomendation

def plant_recomendation_using_TH(Temperature,Humidity,dataset):

for i in range(len(dataset)):

if Temperature>=dataset.MinTemp[i] and Temperature<=dataset.MaxTemp[i]:

if Humidity>=dataset.Minhumid[i] and Humidity<=dataset.Maxhumid[i]:

print("plant :{}\t\t\t soil :{}".format(dataset.Name[i],dataset.Soil)

#classifying inputs based on user choice to call the respective functions for class

def option_fun(c):

if str(c)=="1":
```

```
name=input("Enter Name of the plant:")

recomendation_using_name(name,dataset)

elif str(c)=="2":

season=input("Enter Name of the season: ")

season=season.lower()

plant_recomendation_using_S(season,dataset) elif str(c)=="3":

label=input("Enter the Plant Name: ")

Temperature=int(input("Enter the Average Temperature of your locality: "))
Humidity=int(input("Enter the Average Humidity of your locality: "))

soil=input("Enter the soil type do you check in: ")

soil=soil.lower()#case conversion of the soil to ease of searching in dataset
Label=label[0].upper()+label[1::].lower()

#string manipulation in accordance

#plant_classification function calling using following parameters
plant_classification_using_LTHS(Label,Temperature,Humidity,soil,dataset)

elif str(c)=="4":

Temperature=int(input("Enter the temperature of your locality: "))
Humidity=int(input("Enter the humidity of your locality: "))
plant_recomendation_using_TH(Temperature,Humidity,dataset)

else:

print("Enter a valid entry")

#asking for plant category like fruit,vegetable,herb,flower n=input("Type of plant:")
n=n.lower()
```

```
#explaining the user about which type of classification are behind the options 1,2, 3,4

print("1:Recomendation using name and output is required conditions for that plant")
print("2:Recomendation using season and output is ideal plants for that season")
print("3:classification using name,temperature,humidity,soil and output is either
yes/no")

print("4:Recomendation using Temperature and Humidity,name and soil are outputs")
c=input("choose one of the options (1/2/3/4): ")

#conditonal based dataset training

if n=="herb":

dataset=pd.read_csv(r"D:\final_year_projects\imran\HerbsDataset.csv")

option_fun(c)

elif n=="vegetable":
dataset=pd.read_csv(r"D:\final_year_projects\imran\VegetablesDataset.csv")

option_fun(c)

elif n=="fruit":

dataset=pd.read_csv(r"D:\final_year_projects\imran\FruitsDataset.csv")

option_fun(c)

elif n=="flower":

dataset=pd.read_csv(r"D:\final_year_projects\imran\Flowersdataset.csv")

option_fun(c)

else:

print("valid type of plant")

else:

print("Thank You for your time♡")
```

APPENDIX B

SCREENSHOTS

The seasons considered in this project are

Figure B1.1: Total no.of seasons considered

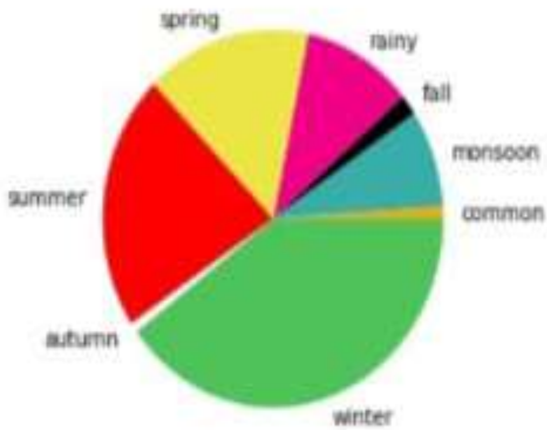

Figure B1.2: Percentage of seasons used in dataset

The different types of soil considered are shown below.

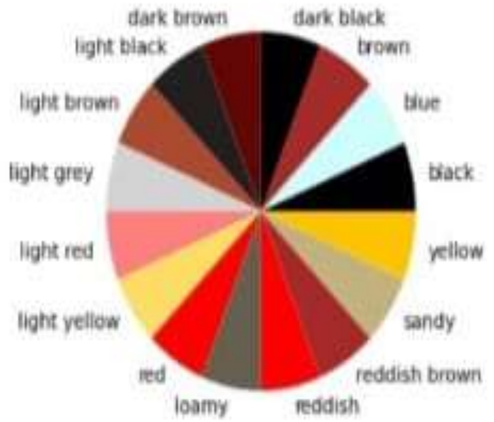

Figure B1.3: Types of soils

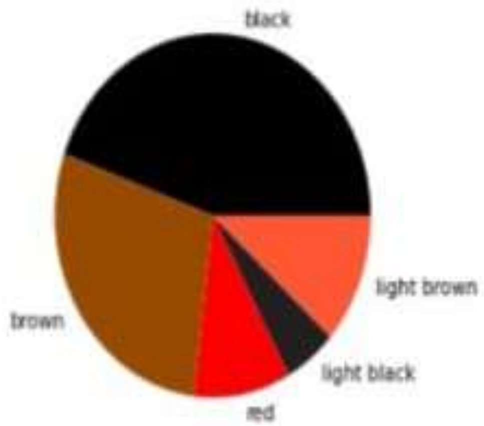

Figure B1.4: Soils Percentage Considered in Dataset

Here few soil types are negligible because few soils are only suitable for few plants growth.

APPENDIX C

STUDENT CONTRIBUTION

No.	ACTIVITY	18K61A05G7	18K61A05D4	18K61A05E0	18K61A05H0
1	Title Conformation	✓	✓	✓	✓
2	Literature Survey	✓	✓	✓	✓
3	Problem Formulation	✓	✓	✓	✓
4	Requirement Gathering	✓	✓	✓	✓
5	Designing	✓	✓	✓	
6	Implementation	✓	✓	✓	
7	Documentation	✓	✓	✓	✓

APPENDIX D

PO, PSO, PEO, AND CO RELEVANCE WITH PROJECT

CO-PO MAPPING SHEET

COURSE OUTCOMES

OUTCOME NO	DESCRIPTION
CO1	Develop problem formation and design skills for engineering and real-world problems.
CO2	Collect and Generate ideas through literature survey on current research areas which help to analyse and present to impart knowledge in different fields.
CO3	Import knowledge on software & hardware to meet industry perspective needs and standards.
CO4	Create interest to carry out research on innovative ideas as a lifelong learning.
CO5	Ability to work with team, and enrich presentation and communication skills.
CO6	Create a platform that makes students employable.

SUMMARY OF CO MAPPING TO PROGRAM OUTCOMES

COs/POs	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO10	PO11	PO12	PSO 1	PSO 2
CO1	3	0	0	1	0	1	1	1	3	3	0	1	0	0
CO2	3	3	0	0	0	2	0	0	3	2	0	0	0	0
CO3	2	0	1	1	3	0	0	0	3	2	0	0	0	0
CO4	3	0	0	3	3	0	3	1	3	3	1	1	0	0
CO5	2	0	0	0	2	0	0	0	3	3	0	3	0	0
CO6	2	1	0	0	3	1	0	3	3	2	2	2	0	0
Overall Course	3	1	1	1	2	1	1	0	3	2	1	1	0	0

PROGRAM OUTCOMES (POs)

POs	PROGRAM OUTCOMES	RELEVANCE
PO1	**Engineering Knowledge:** Apply knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.	This project needs a mathematics and Computer Science and Engineering specialization background to perform calculations in the classification task.
PO2	**Problem Analysis:** Identity, formulates, research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.	For this project ruderous literature survey is conducted to analyze the existing systems problems.
PO3	**Design/ Development of Solutions:** Design solutions for complex engineering problems and design system components or processes that meet specified needs with appropriate consideration for public health and safety, cultural, societal, and environmental considerations.	Once the formulation of the problem has been completed, the design of the solution relevant to the problem is created to meet the needs of the problem in all aspects.
PO4	**Conduct investigations of complex problems:** Using research-based knowledge and research methods including design of experiments, analysis, and interpretation of data, and synthesis of the information to provide valid conclusions.	Referred to similar kinds of experiments to gain the knowledge of fixing parameters and framing the conclusions.
PO5	**Modern Tool Usage:** Create, select and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.	Recent methods like Jupiter notebook have been used to solve the stated problem
PO6	**The Engineer and Society:** Apply to reason informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to professional engineering practice.	This problem provides a solution to the people without depletion of any cultural, social, health, safety, and legal issues.

PO7	**Environment and Sustainability:** Understand the impact of professional engineering solutions in societal and environmental contexts and demonstrate knowledge of and need for sustainable development.	This Project doesn't deteriorate any sort of environmental issues.
PO8	**Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of engineering practice.	This project has been executed by following proper ethics as stated in the engineering practice.
PO9	**Individual and Team Work:** Function effectively as an individual, and as a member or leader in diverse teams and multidisciplinary settings.	This project is carried out with collective teamwork by making the entire project into proper segments.
PO10	**Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations and give and receive clear instructions.	Complete information related to the project has been documented for clear understanding.
PO11	**Life-long Learning:** Recognize the need for and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.	----
PO12	**Project Management and Finance:** Demonstrate knowledge and understanding of engineering and management principles and apply these to one's work, as a member and leader in a team, to manage projects and in multidisciplinary environments.	This work can be enhanced to a larger extent concerning time and other factors.

PROGRAM SPECIFIC OUTCOME (PSOs)

PSOs	Program Specific Outcome	Relevance
PSO1	**Mobile & Web Application Development:** Ability to develop mobile & web applications using J2EE, Android, and J2ME.	-----
PSO2	**Cloud Services:** To deploy virtualized and cloud-based services in the organization.	-----

PROGRAMME EDUCATIONAL OBJECTIVES (PEOs)

PEOs	Programme Educational Objectives	RELEVANCE
PEO 1	Graduates will be able to analyze, design, and develop advanced computer applications to provide a solution to real-world problems.	To get the project executed, all the team have done analysis and research-oriented surveys to frame the solution and identify the limitations.
PEO 2	Graduates are well-trained, confident, research-oriented, and industry-ready professionals who are intellectual, ethical, and socially committed.	As implemented the problem with Deep Learning, the latest trend that leads to being well accustomed to the recent technological standards as per industry requirement.
PEO 3	Graduates will have the technical, communication skills and character that will prepare them for technical and leadership roles.	After completing the project successfully, all the team members can be able to reach a satisfactory level in explaining technical aspects with effective communication.

COURSE OUTCOME (COs)

COs	Course Outcome	POs, PSOs and PEOs Mapped
CO1	Develop problem formation and design skills for engineering and real-world problems	PO1, PO2, PO3, PSO2
CO2	Collect and Generate ideas through literature surveys on current research areas which help to analyze and	PO2, PO3, PO5, PO6
CO3	Import knowledge on software & hardware to meet industry perspective needs and standards.	PO5, PO7, PO8, PO9
CO4	Create interest to research innovative ideas as lifelong learning.	PO11
CO5	Ability to work with a team, and enrich presentation and communication skills.	PO10
CO6	Create a platform that makes students employable.	PO5, PO9, PO11, PO12, PSO2

RELEVANCE TO POs

CO	PO	PI	Relevance
CO1	PO1	1.2.1	Apply different statistics and numerical techniques to solve the problem.
	PO4	4.4.2	Understand the problem and applied the proper algorithm.
	PO6	6.4.1	This is challenged state to assess societal, safety and legal issues.
	PO7	7.3.1	Identified the risks/impacts in the life-cycle of an product and activity.
	PO8	8.3.1	Identified situations of unethical professional conduct and propose ethical alternatives.
	PO9	9.5.2	This work is carried out by all the team members.
	PO10	10.4.2	Produced the work in well-structured form.
	PO12	12.5.2	This work can be enhanced to larger extent with respect to the time and other factors.
CO2	PO1	1.6.1	Uses the engineering fundamentals to complete the work.
	0PO2	2.6.4	Compared and select alternative solution/methods to select the best methods.
	PO6	6.4.1	Interpret legislation, regulations, codes, and standards relevant to your discipline and explain its contribution to the protection of the public.
	PO9	9.5.2	Treat other team members respectfully.
	PO10	10.4.2	Produced the work in well-structured form.
	PO1	1.2.1	Applied the knowledge of discrete structures, linear algebra, statistics and numerical techniques to solve problems.

CO3	PO3	3.6.2	Ability to produce a variety of potential design solutions suited to meet functional requirements.
	PO4	4.4.3	Ability to choose appropriate hardware/software tools to conduct the experiment.
	PO5	5.5.1	Identify the strengths and limitations of tools for (i) acquiring information, (ii) modelling and simulating, (iii) monitoring system performance, and (iv) creating engineering designs.
	PO9	9.4.2	Implement the norms of practice (e.g. rules, roles, charters, agendas, etc.) of effective team work, to accomplish a goal.
	PO10	10.4.1	Read, understand and interpret technical and nontechnical information.
CO4	PO1	1.5.1	Apply laws of natural science to an engineering problem.
	PO4	4.6.2	Critically analyse data for trends and correlations, stating possible errors and limitations.
	PO5	5.6.2	Verify the credibility of results from tool use with reference to the accuracy and limitations, and the assumptions inherent in their use.
	PO7	7.4.1	Describe management techniques for sustainable development
	PO8	8.4.2	Examine and apply moral & ethical principles to known case studies
	PO9	9.5.2	Treat other team members respectfully
	PO10	10.5.2	Deliver effective oral presentations to technical and nontechnical audiences
	PO11	11.4.2	Analyze different forms of financial statements to evaluate the financial status of an engineering project

CO5	PO1	1.6.1	Apply engineering fundamentals
	PO5	5.5.2	Demonstrate proficiency in using discipline specific tools.
	PO9	9.5.3	Listen to other members ure in difficult situations
	PO10	10.5.1	Listen to and comprehend information, instructions, and viewpoints of others
	PO12	12.6.2	Analyze sourced technical and popular information for feasibility, viability, sustainability, etc.
CO6	PO1	1.7.1	Apply theory and principles of computer science engineering to solve an engineering problem.
	PO2	2.6.2	Identifies functionalities and computing resources.
	PO5	5.6.1	Discuss limitations and validate tools, techniques and resources
	PO6	6.3.1	Identify and describe various engineering roles; particularly as pertains to protection of the public and public interest at global, regional and local level.
	PO8	8.3.1	Identify situations of unethical professional conduct and propose ethical alternatives.
	PO9	9.5.1	Demonstrate effective communication, problem solving, conflict resolution and leadership skills
	PO10	10.5.1	Listen to and comprehend information, instructions, and viewpoints of others
	PO11	11.6.1	Identify the tasks required to complete an engineering activity, and the resources required to complete the tasks.
	PO12	12.6.1	Source and comprehend technical literature and other credible sources of information.

CO	PSO	Relevance
CO1	-	-
CO2	-	-
CO3	-	-
CO4	-	-
CO5	-	-
CO6	-	-