CROP YEILD PREDICTION AND CROP RECOMMENDATION BASED ON MACHINE LEARNING

Submitted in partial fulfillment of the requirements
for the award of
Bachelor of Engineering degree in Computer Science and Engineering

by

Budha Pretesh (38110089)
Ch Sai Teja (38110101)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited with Grade "A" by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI - 600 119

MARCH – 2022

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with "A" grade by NAAC

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai - 600119

www.sathyabama.ac.in

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **Budha Pretesh (38110089)** and **Ch Sai Teja (38110101)** who carried out the project entitled "**CROP YEILD PREDICTION AND CROP RECOMMENDATION BASED ON MACHINE LEARNING"** under my supervision from November 2021 to March 2022.

Internal Guide N

Head of the Department

Submitted for Viva voce Examination held on _____

Internal Examiner **External Examiner**

DECLARATION

I **Budha Pretesh (Reg No:38110101) and Ch Sai Teja(Reg No: 38110101)** hereby declare that the Project Report entitled "**Crop Yield Prediction & Recommendation Based on machine Learning**" done by us under the guidance of **Mr.Murari Devakannan Kamalesh** is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in 2018-2022.

DATE:

PLACE: **SIGNATUREOF THE CANDIDATE**

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T.Sasikala M.E.,Ph.D.**, **Dean**, School of Computing **Dr.S.Vigneshwari M.E., Ph.D.** and **Dr.L.Lakshmanan M.E., Ph.D. ,** Heads of the Department of Computer Science and Engineering for providing us necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and a deep sense of gratitude to my Project Guide **Mr.Murari Devakannan Kamalesh** for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

Agriculture growth mainly depends on diverse soil parameters, namely Nitrogen, Phosphorus, Potassium, Crop rotation, Soil moisture, pH, surface temperature and weather aspects like temperature, rainfall, etc. Technology will prove to be beneficial to agriculture which will increase crop productivity resulting in better yields to the farmer. The proposed project provides a solution for Smart Agriculture by monitoring the agricultural field which can assist the farmers in increasing productivity to a great extent. This work presents a system, in a form of a website, which uses Machine Learning techniques in order to predict the most profitable crop in the current weather and soil conditions. This system can also help in predicting the yield of the crop using weather parameter, soil parameter and historic crop yield. Thus, the project develops a system by integrating data from various sources, data analytics, prediction analysis which can improve crop yield productivity and increase the profit margins of farmer helping them over a longer run.

TABLE OF CONTENTS

xxxxxxxxxxxxxxx

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS	EXPANSION
ML	Machine Learning
NB	Naive Bayes
KNN	Kth nearest neighbour

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

Crop yield prediction is one of the challenging tasks in agriculture. It plays an essential role in decision making at global, regional, and field levels. The prediction of crop yield is based on soil, meteorological, environmental, and crop parameters. Decision support models are broadly used to extract significant crop features for prediction. Precision agriculture focuses on monitoring (sensing technologies), management information systems, variable rate technologies, and responses to inter- and intravariability in cropping systems. The benefits of precision agriculture involve increasing crop yield and crop quality, while reducing the environmental impact.

Crop yield simulations help to understand the cumulative effects of water and nutrient deficiencies, pests, diseases, the impact of crop yield variability, and other field conditions over the growing season.

1.2 MACHINE LEARNING

Machine learning could be a subfield of computer science (AI). The goal of machine learning typically is to know the structure information of knowledge of information and match that data into models which will be understood and used by folks. Although machine learning could be a field inside technology, it differs from ancient process approaches.

In ancient computing, algorithms are sets of expressly programmed directions employed by computers to calculate or downside solve. Machine learning algorithms instead give computers to coach on knowledge inputs and use applied math analysis so as to output values that fall inside a particular vary. thanks to this, machine learning facilitates computers in building models from sample knowledge so as to modify decision-making processes supported knowledge inputs.

1.2 OBJECTIVE

The main objective this paper introduces a ML model that classify and predict crop yeild and predict crop by utilizing supervised ML algorithms. Thus, the proposed approach offers a solution to predict performance efficiently and accurately by comparing several ML model.

CHAPTER 2

LITERATURE SURVEY:

[1]Title: Crop Yield Prediction using Machine Learning Algorithm

D.Jayanarayana Reddy; M. Rudra Kumar

Agribusiness is the foundation of India's economy, with in excess of 50% of the populace occupied with cultivating. Environmental change,

environmental change and other ecological variables altogether affect farming wellbeing. Machine learning (ML) assumes a significant part as it is an apparatus for Crop Yield Prediction (CYP) independent direction.

[2]Title: **Crop Yield Prediction in Precision Agriculture**

Prof. Dr. Miklós Neményi

The prediction of crop yield is based on soil, meteorological, environmental, and crop parameters. Decision support models are broadly used to extract significant crop features for prediction. Precision agriculture focuses on monitoring (sensing technologies), management systems, variable rate technologies, and responses to inter- and intervisibility cropping systems.

[3]Title: **Crop Recommender System Using Machine Learning Approach**

Shilpa mangesh pande; prem kumar ramesh; anmol; b. R aishwarya; karuna rohilla

There is no question that farming and enterprises are connected to the jobs of country Indians. This is one of the principle justifications for why peripheral ranchers in India end it all

[4]Title: **AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms**

ZeelDoshi; Subhash Nadkarni; Rashi Agrawal

horticulture is a significant commitment to the Indian economy. The huge number of individuals living in India relies upon how they live in

horticulture. Numerous Indian ranchers accept that they can pick plants to plant at a given time.

[5]Title: A Review on Data Mining Techniques for Fertilizer Recommendation, 2018

Authors: Jignasha M. Jethva, Nikhil Gondaliya, Vinita Shah

At the point when the dirt is insufficient in supplements, add compost to decrease it. A typical issue in farming is excrement determination and fertilizer expansion. Extreme development or absence of manure can harm vegetation and diminish efficiency. This record sums up the different techniques for removing information used to develop a bunch of modern manure soils.

[6]Title: A Survey on Data Mining Techniques in Agriculture, 2015

Author: M.C.S.Geetha

Horticulture is important for the economies of emerging nations, particularly India. Mining assumes a significant part in decision-production in numerous spaces of farming. It inspects the job of data mining in the farming area and work corresponding to a couple of creators in the rural area. It likewise checks out various methods of acquiring data to resolve numerous agrarian issues. This paper integrates the work of several authors in a single place so it is valuable for specialists to get data of current situation of data mining systems and applications in context to farming field.

[7]Title: AgroNutri Android Application, 2016

Authors: S. Srija, R. Geetha Chanda, S.Lavanya, Dr. M. Kalpana Ph.D

This paper communicates the idea regarding the making of AgroNutri an android application that helps in conveying the harvest particular fertilizer amount to be applied. The future scope of the AgroNutri is that GPRS can be included so that according to location nutrients are suggested. Further this application would be incorporated as a piece of the accuracy agriculture wherein sensors can be utilized to discover the measure of NPK present in the dirt and that sum can be deducted from the suggestion and giving us the exact measure of supplements to be added.

[8]Title: Machine Learning: Applications in Indian Agriculture, 2016

Authors: Karandeep Kaur

Agribusiness was an area that was contrary with innovation and its turn of events. Indian ranchers should comply with the common principles. AI is the fundamental thought utilized in all data sources and results. He has utilized his capacity in basic science and programming. Mechanical preparing figures have incredibly worked on the craft of AI and incorporate sensor-based parts utilized in coordinated agribusiness. This paper analyses the different employments of AI in horticultural fields. It gives experiences into the issues looked by ranchers and how to settle them utilizing this technique.

Existing System

- Even now, Agriculture supports about 58% of total population, which is 75% at the time of independence i.e., a drop of 17%.
- A good amount of people in villages are leaving the agriculture and adopting other professions due to poor yields and returns.

In a recent study, about 76% of farmers want to give up farming as there is no market and amount of production

Proposed System

- Large amount of datasets are given containing information about types of crops, yield, soil types, seasons etc.
- Providing the user to select based on which field he/she wants to perform the analysis.
- Analyzing the data present with us based on the provided user requirement.
- Showing results with at most available possibilities which will enhance the chances of yield in the future.

Major Prerequisites:

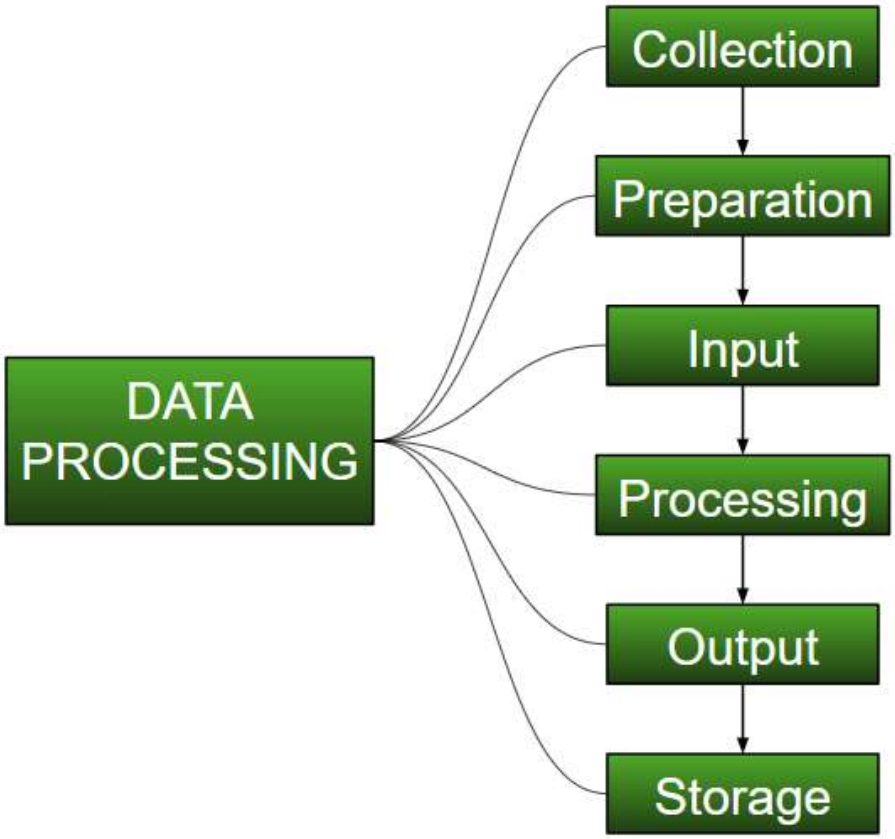
- Datasets containing information about various types of crops.
- Data related to amount of yield produced by variety of crops.
- An efficient algorithm for analysis of yield.

Implementation:

We proposed as an alternative to the user-based neighborhood approach. We first consider the dimensions of the input and output of the neural network. In order to maximize the amount of training data we can feed to the network, we consider a training example to be a user profile (i.e. a

row from the user-item matrix R) with one rating withheld. The loss of the network on that training example must be computed with respect to the single withheld rating. The consequence of this is that each individual rating in the training set corresponds to a training example, rather than each user. As we are interested in what is essentially a regression, we choose to use root mean squared error (RMSE) with respect to known ratings as our loss function. Compared to the mean absolute error, root mean squared error more heavily penalizes predictions which are further off. We reason that this is good in the context of recommender system because predicting a high rating for an item the user did not enjoy significantly impacts the quality of the recommendations. On the other hand, smaller errors in prediction likely result in recommendations that are still useful—perhaps the regression is not exactly correct, but at least the highest predicted rating are likely to be relevant to the user.

Data Processing is a task of converting data from a given form to a much more usable and desired form i.e. making it more meaningful and informative. Using Machine Learning algorithms, mathematical modeling and statistical knowledge, this entire process can be automated. The output of this complete process can be in any desired form like graphs, videos, charts, tables, images and many more, depending on the task we are performing and the requirements of the machine. This might seem to be simple but when it comes to really big organizations like Twitter, Facebook, Administrative bodies like Paliament, UNESCO and health sector organizations, this entire process needs to be performed in a very structured manner.



Collection:

The most crucial step when starting with ML is to have data of good quality and accuracy. Data can be collected from any authenticated source like data.gov.in, Kaggle or UCI dataset repository.For example, while preparing for a competitive exam, students study from the best study material that they can access so that they learn the best to obtain the best results. In the same way, high-quality and accurate data will make the learning process of the model easier and better and at the time of testing, the model would yield state of the art results.

A huge amount of capital, time and resources are consumed in collecting data. Organizations or researchers have to decide what kind of data they need to execute their tasks or research.

Example: Working on the Facial Expression Recognizer, needs a large

number of images having a variety of human expressions. Good data ensures that the results of the model are valid and can be trusted upon.

Preparation:

The collected data can be in a raw form which can't be directly fed to the machine. So, this is a process of collecting datasets from different sources, analyzing these datasets and then constructing a new dataset for further processing and exploration. This preparation can be performed either manually or from the automatic approach. Data can also be prepared in numeric forms also which would fasten the model's learning.

Example: An image can be converted to a matrix of N X N dimensions, the value of each cell will indicate image pixel.

Processing:

This is the stage where algorithms and ML techniques are required to perform the instructions provided over a large volume of data with accuracy and optimal computation.

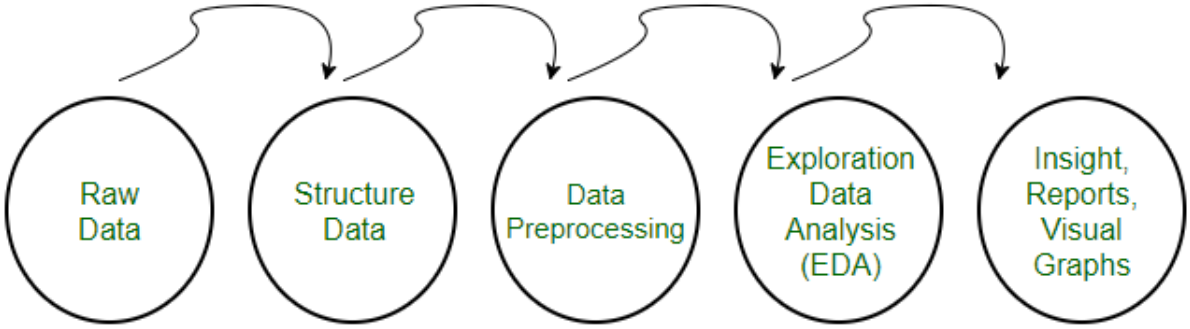
Output:

In this stage, results are procured by the machine in a meaningful manner which can be inferred easily by the user. Output can be in the form of reports, graphs, videos, etc

Data Preprocessing for Machine learning in Python

• Pre-processing refers to the transformations applied to our data before feeding it to the algorithm.

• Data Preprocessing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.



Need of Data Preprocessing

• For achieving better results from the applied model in Machine Learning projects the format of the data has to be in a proper manner. Some specified Machine Learning model needs information in a specified format, for example, Random Forest algorithm does not support null values, therefore to execute random forest algorithm null values have to be managed from the original raw data set.

• Another aspect is that data set should be formatted in such a way that more than one Machine Learning and Deep Learning algorithms are executed in one data set, and best out of them is chosen.

Rescale Data

- When our data is comprised of attributes with varying scales, many machine learning algorithms can benefit from rescaling the attributes to all have the same scale.

- This is useful for optimization algorithms in used in the core of machine learning algorithms like gradient descent.
- It is also useful for algorithms that weight inputs like regression and neural networks and algorithms that use distance measures like K-Nearest Neighbors.
- We can rescale your data using scikit-learn using the MinMaxScaler class.

Binarize Data (Make Binary)

- We can transform our data using a binary threshold. All values above the threshold are marked 1 and all equal to or below are marked as 0.
- This is called binarizing your data or threshold your data. It can be useful when you have probabilities that you want to make crisp values. It is also useful when feature engineering and you want to add new features that indicate something meaningful.
- We can create new binary attributes in Python using scikit-learn with the Binarizer class.

Standardize Data

- Standardization is a useful technique to transform attributes with a Gaussian distribution and differing means and standard deviations to a standard Gaussian distribution with a mean of 0 and a standard deviation of 1.
- We can standardize data using scikit-learn with the StandardScaler class.

Data Cleansing

Introduction:

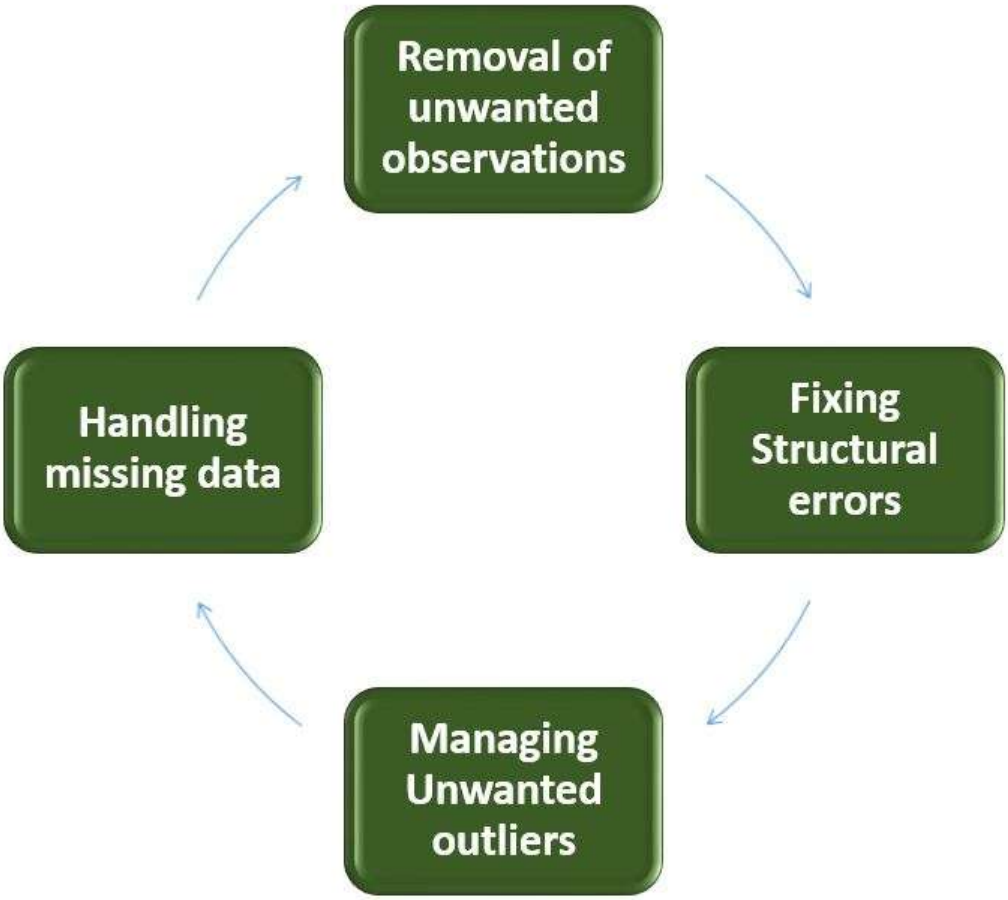
Data cleaning is one of the important parts of machine learning. It plays a significant part in building a model. Data Cleaning is one of those things that everyone does but no one really talks about. It surely isn't the fanciest part of machine learning and at the same time, there aren't any hidden tricks or secrets to uncover. However, proper data cleaning can make or break your project. Professional data scientists usually spend a very large portion of their time on this step.

Because of the belief that, "Better data beats fancier algorithms".

If we have a well-cleaned dataset, we can get desired results even with a very simple algorithm, which can prove very beneficial at times.

Obviously, different types of data will require different types of cleaning. However, this systematic approach can always serve as a good starting point.

Steps involved in Data Cleaning



1. Removal of unwanted observations

This includes deleting duplicate/ redundant or irrelevant values from your dataset. Duplicate observations most frequently arise during data collection and irrelevant observations are those that don't actually fit the specific problem that you're trying to solve.

- Redundant observations alter the efficiency by a great extent as the data repeats and may add towards the correct side or towards the incorrect side, thereby producing unfaithful results.
- Irrelevant observations are any type of data that is of no use to us and can be removed directly.

2. Fixing Structural errors

The errors that arise during measurement transfer of data or other similar situations are called structural errors. Structural errors include typos in the name of features, same attribute with different name, mislabeled classes, i.e. separate classes that should really be the same or inconsistent capitalization.

- For example, the model will treat America and america as different classes or values, though they represent the same value or red, yellow and red-yellow as different classes or attributes, though one class can be included in other two classes. So, these are some structural errors that make our model inefficient and gives poor quality results.

3. Managing Unwanted outliers

Outliers can cause problems with certain types of models. For example, linear regression models are less robust to outliers than decision tree models. Generally, we should not remove outliers until we have a legitimate reason to remove them. Sometimes, removing them improves performance, sometimes not. So, one must have a good reason to remove the outlier, such as suspicious measurements that are unlikely to be the part of real data.

4. Handling missing data

Missing data is a deceptively tricky issue in machine learning. We cannot just ignore or remove the missing observation. They must be handled

carefully as they can be an indication of something important. The two most common ways to deal with missing data are:

1. Dropping observations with missing values.

Dropping missing values is sub-optimal because when you drop observations, you drop information.

- The fact that the value was missing may be informative in itself.
- Plus, in the real world, you often need to make predictions on new data even if some of the features are missing!

2. Imputing the missing values from past observations.

Imputing missing values is sub-optimal because the value was originally missing but you filled it in, which always leads to a loss in information, no matter how sophisticated your imputation method is.

- Again, "missingness" is almost always informative in itself, and you should tell your algorithm if a value was missing.
- Even if you build a model to impute your values, you're not adding any real information. You're just reinforcing the patterns already provided by other features.
- Both of these approaches are sub-optimal because dropping an observation means dropping information, thereby reducing data and imputing values also is sub-optimal as we fil the values that were not present in the actual dataset, which leads to a loss of information.
- Missing data is like missing a puzzle piece. If you drop it, that's like pretending the puzzle slot isn't there. If you impute it, that's like trying to squeeze in a piece from somewhere else in the puzzle.

- So, missing data is always informative and indication of something important. And we must aware our algorithm of missing data by flagging it. By using this technique of flagging and filling, you are essentially allowing the algorithm to estimate the optimal constant for missingness, instead of just filling it in with the mean.

CHAPTER 3

METHODOLOGY

3.1 EXISTING SYSTEM

Even now, Agriculture supports about 58% of total population, which is 75% at the time of independence i.e., a drop of 17%.A good amount of people in villages are leaving the agriculture and adopting other professions due to poor yields and returns.In a recent study, about 76% of farmers want to give up farming as there is no market and amount of production

3.2 PROPOSED SYSTEM

In the propesed system Machine learning algorithms are used for the prediction and recommendation. Large amount of datasets are given containing information about types of crops, yield, soil types, seasons etc. Providing the user to select based on which field he/she wants to perform the analysis.Analyzing the data present with us based on the provided user requirement.Showing results with at most available possibilities which will enhance the chances of yield in the future.

3.3 Algorithms USed

Decision Tree Classifier

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, , but mostly it is preferred for solving Classification problems.

The decision tree classifier creates the classification model by building a decision tree.

Each node in the tree specifies a test on an attribute, each branch descending from that node corresponds to one of the possible values for that attribute.

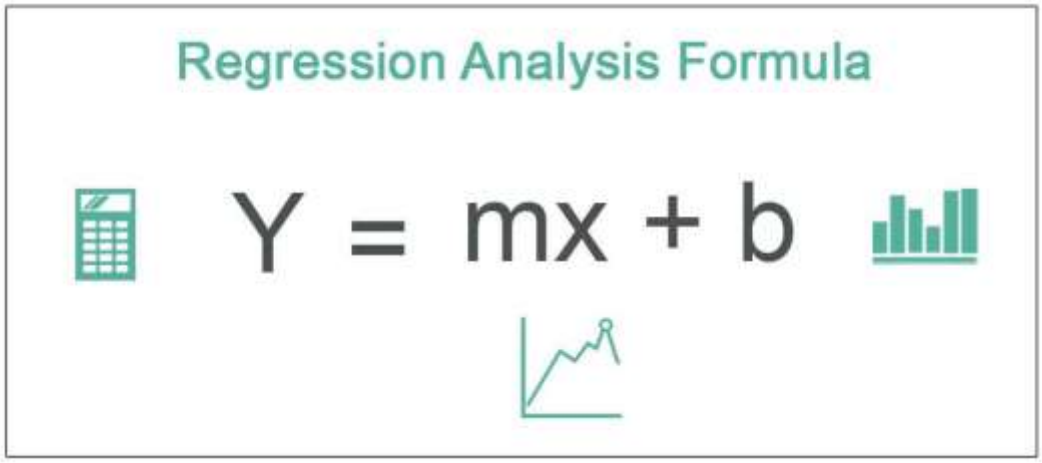
$$Gini = 1 - \sum_{i=1}^{C} (p_i)^2$$

Linear Regression

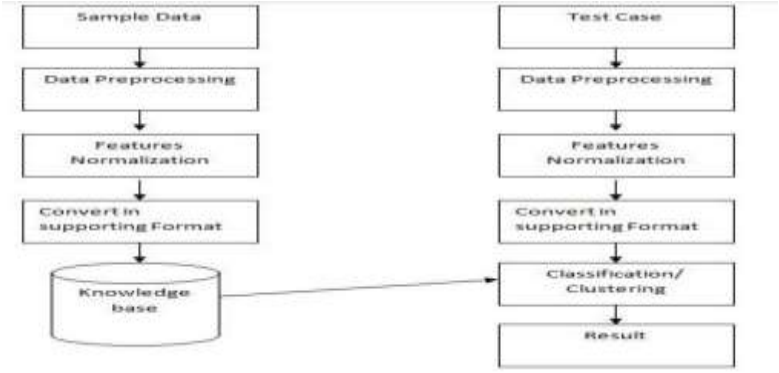
Linear Regression is the process of finding a line that best fits the data points available on the plot, so that we can use it to predict output values for inputs that are not present in the data set we have, with the belief that those outputs would fall on the line.

The Statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variable.

Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables, hence called as linear regression.



3.4 SYSTEM ARCHITECTURE



3.5 SYSTEM REQUIREMENTS

HARDWARE REQUIREMENTS

Processor	:	Intel i3 and above
RAM	:	4GB and Higher
Hard Disk	:	500GB: Minimum

SOFTWARE REQUIREMENTS:

Programming Language	: Python
IDE	: PyCharm
UML Design	: Start UML
Tools	: PIP
WebServer	: Tomcat
Framework	: DJango, Google Maps API
Database	: SQLite

3.6 MODULES

Data Collection

Preprocessing Step

Feature Extraction

Data Prediction

Data Collection

Data is composed from a different source and optimized for data sets. And the data is used to evaluate descriptively. Several abstract online outlets, like Kaggle, Google weather forestation and data government, provide the data for up to 10years in series. The data sets such as soil nature, climatic conditions and seed data are used for the crop prediction and better crop yields.

Preprocessing Step

Preprocessing the data is considered as a significant step machine learning phase. Preprocessing involves adding the missing values, the correct set of data, and extracting the functionality. Data set form is important to the process of analysis. The data collected in this step will induced in Google Colab platform in the form of python programming in order to get the desired output.

Feature Extraction

Extraction of the features would reduce the data size involved to characterize a wide collection of data. The characteristics of soil, crop and weather collected from the pretreatment process establish the final training data collection. This approach selects the features based on the correlation matrix i.e. the features that has more correlation value is selected as an important predictive function for yield.

Data Prediction

In advance to this step there need to split the data into train dataset and test dataset. By applying the random forest the data is trained with available input and output data. In the test phase, the data are tested if the accuracy of e model is satisfied. Then the new data is predicted by machine learning module.

3.7 UML DIAGRAMS

UML is simply anther graphical representation of a common semanticmodel. UML provides a comprehensive notation for the full lifecycle of object-oriented development.

ADVANTAGES

To represent complete systems (instead of only the software portion) using object oriented concepts

- To establish an explicit coupling between concepts and executable code
- To take into account the scaling factors that are inherent to complex and critical systems
- To creating a modeling language usable by both humans and machines

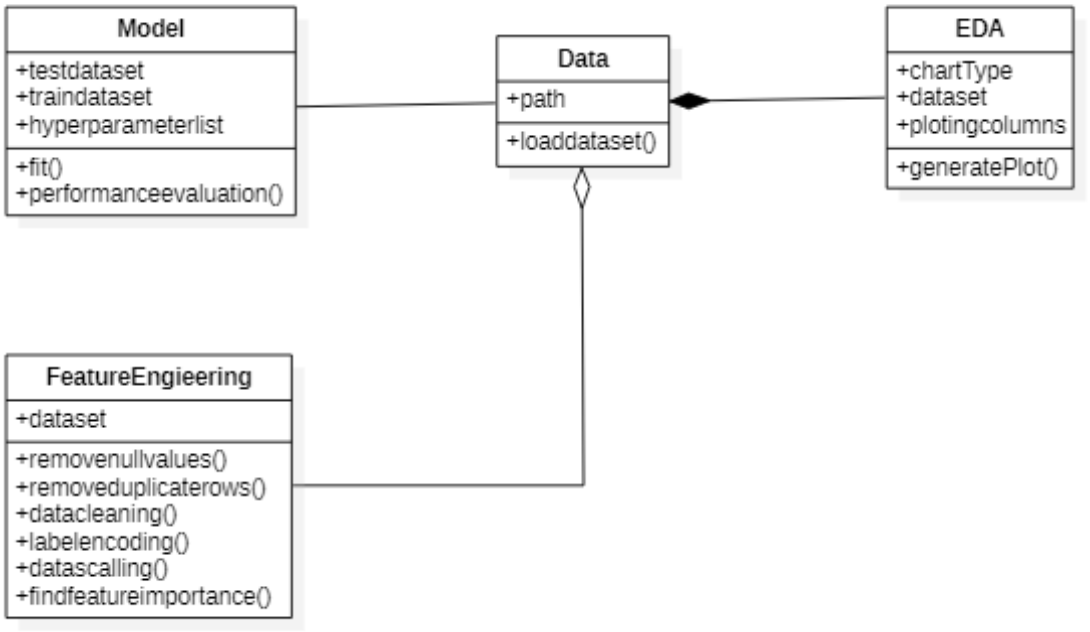
UML defines several models for representing systems

- The class model captures the static structure
- The state model expresses the dynamic behavior of objects
- The use case model describes the requirements of the user
- The interaction model represents the scenarios and messages flows
- The implementation model shows the work units

- The deployment model provides details that pertain to process allocation

CLASS DIAGRAM:

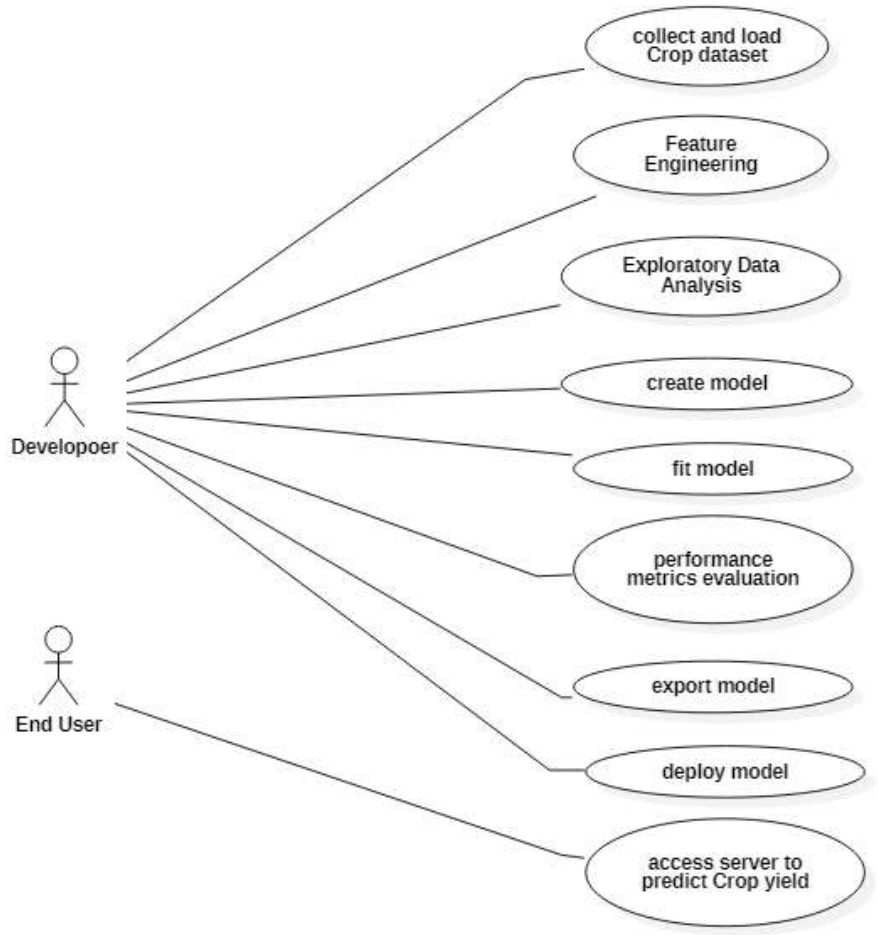
In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.



- **USECASE DIAGRAM**

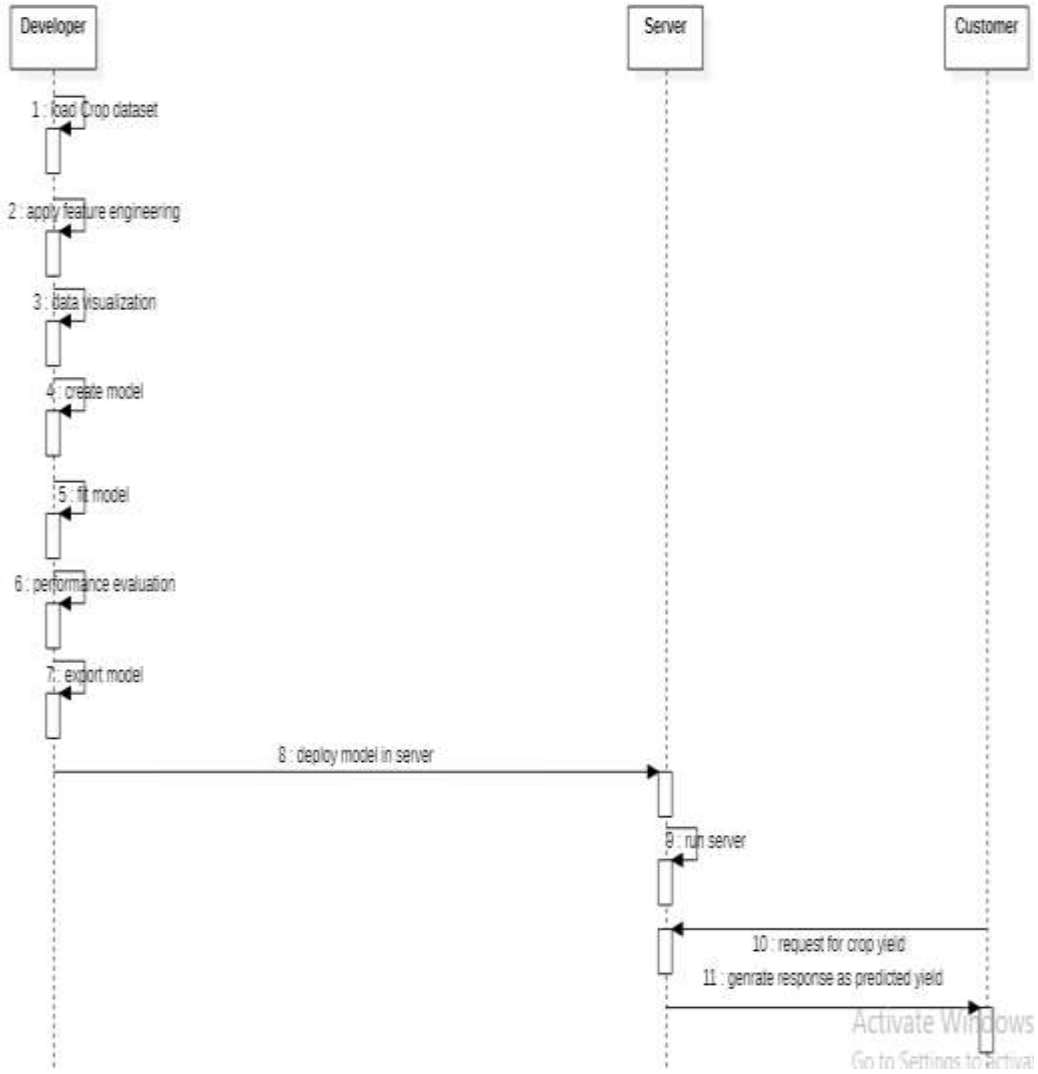
Use case diagrams overview the usage requirement for system. They are useful for presentations to management and/or project stakeholders, but for actual development you will find that use cases

provide significantly more value because they describe "the meant" of the actual requirements. A use case describes a sequence of action that provides something of measurable value to an action and is drawn as a horizontal ellipse.



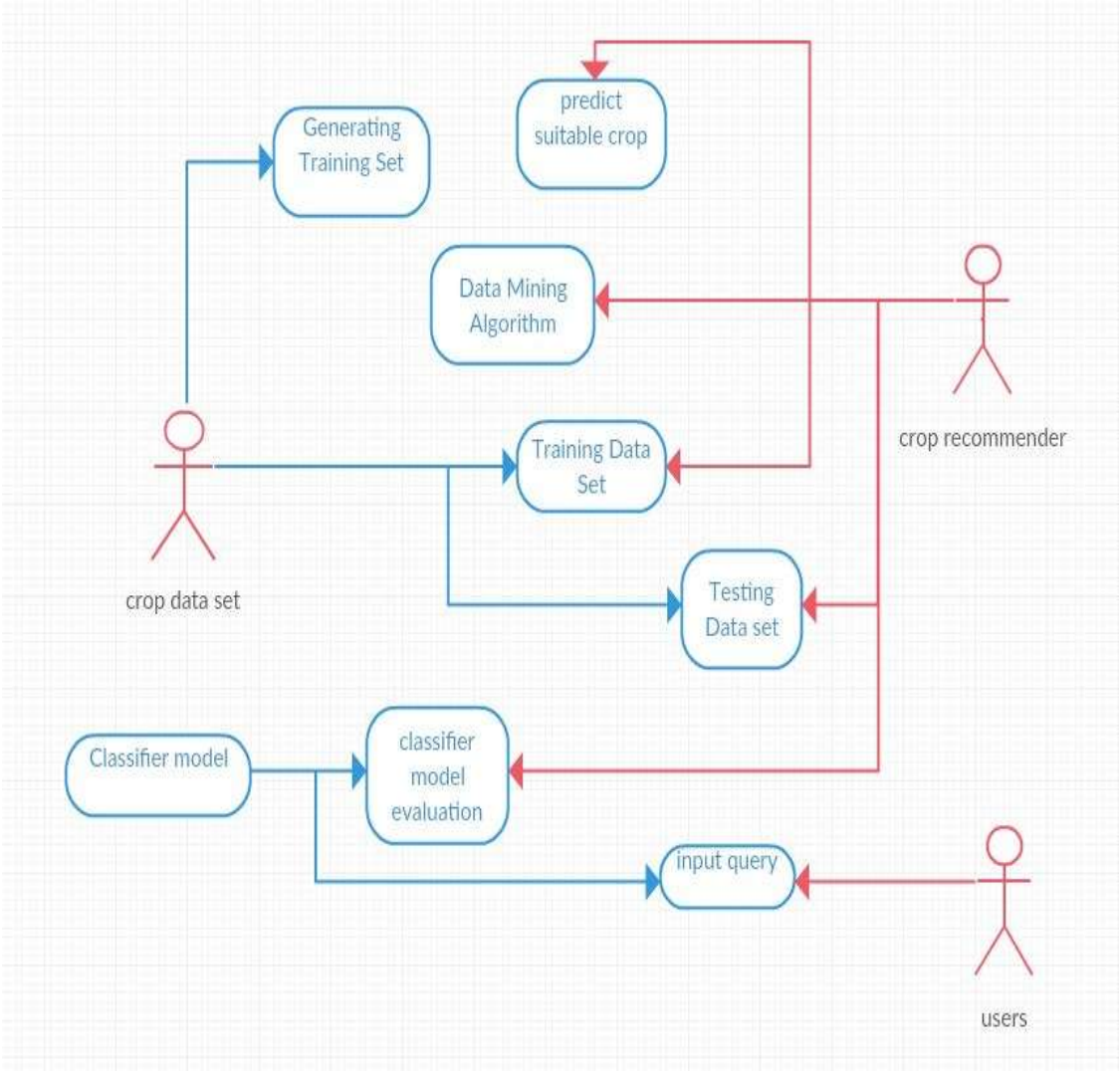
- **SEQUENCE DIAGRAM**

Sequence diagram model the flow of logic within your system in a visual manner, enabling you both to document and validate your logic, and commonly used for both analysis and design purpose. Sequence diagram are the most popular UML artifact for dynamic modeling, which focuses on identifying the behavior within your system.



- **ACTIVITY DIAGRAM**

Activity diagram are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. The activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. Activity diagram consist of Initial node, activity final node and activities in between.



3.8 LANGUAGES USED

PYTHON

Python is an interpreter, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures,

combined with dynamic typing and dynamic binding; make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed. Often, programmers fall in love with Python because of the increased productivity it provides. Since there is no compilation step, the edit-test-debug cycle is incredibly fast. Debugging Python programs is easy: a bug or bad input will never cause a segmentation fault. Instead, when the interpreter discovers an error, it raises an exception. When the program doesn't catch the exception, the interpreter prints a stack trace. A source level debugger allows inspection of local and global variables, evaluation of arbitrary expressions, setting breakpoints, stepping through the code a line at a time, and so on. The debugger is written in Python itself, testifying to Python's introspective power. On the other hand, often the quickest way to debug a program is to add a few print statements to the source: the fast edit-test-debug cycle makes this simple approach very effective.

DJANGO

Django is a Python-based web framework which allows you to quickly create web application without all of the installation or dependency problems that you normally will find with other frameworks.

When you're building a website, you always need a similar set of components: a way to handle user authentication (signing up, signing in, signing out), a management panel for your website, forms, a way to upload files, etc. Django gives you ready-made components to use.

3.9 Requirement Analysis:

Taking into account the comparative analysis stated in the previous section we could start specifying the requirements that our website should achieve. As a basis, an article on all the different requirements for software development was taken into account during this process. We divide the requirements in 2 types: functional and nonfunctional requirements.

Functional requirements

Functional requirement should include function performed by a specific screen outline work-flows performed by the system and other business or compliance requirement the system must meet.

Functional requirements specify which output file should be produced from the given file they describe the relationship between the input and output of the system, for each functional requirement a detailed description of all data inputs and their source and the range of valid inputs must be specified.

The functional specification describes what the system must do, how the system does it is described in the design specification.

If a user requirement specification was written, all requirements outlined in the user requirements specifications should be addressed in the functional requirements.

- The user should be able to register and manage his appointments online at any time.
- Database has to store all the information efficiently without any information loss.
- The user shall be able to search for the doctors by specialty, name, working time and/or gender.
- The user can change his profile info at any time
- hospital can manage all appointments made with him on his account

Nonfunctional requirement

Describe user-visible aspects of the system that are not directly related with the functional behavior of the system. Non-Functional requirements include quantitative constraints, such as response time (i.e. how fast the system reacts to user commands.) or accuracy (.e. how precise are the systems numerical answers.).

- Portability
- Reliability
- Usability
- Time Constraints
- Error messages
- Actions which cannot be undone should ask for confirmation
- Responsive design should be implemented
- Space Constraints
- Performance
- Standards
- Ethics
- Interoperability
- Security
- Privacy

- Scalability.

System Analysis

To provide flexibility to the users, the interfaces have been developed that are accessible through a browser. The GUI'S at the top level have been categorized as

Analysis

Although the scale of this project is relatively small, to produce a professional solution is it imperative that the current problem is understood accurately. However, this task has been made doubly difficult by the lack of support from the company. Thankfully, the Application manager has been kind enough to spare me some of his own time to discuss the problem with me further. Therefore, this chapter is concerning with analyzing the current situation and expectations of the user for this system.

Requirements:

The minimum requirements of the project are listed below:

- Examine the tools and methodologies required to gain an overview of the system requirements for the proposed database.
- Examine suitable database management systems that can be used to implement the proposed database.
- Evaluate appropriate website authoring and web graphic creation tools that can be used to develop web based forms for the proposed database
- Produce and apply suitable criteria for evaluating the solution

FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

The feasibility study investigates the problem and the information needs of the stakeholders. It seeks to determine the resources required to provide an information systems solution, the cost and benefits of such a solution, and the feasibility of such a solution.

The goal of the feasibility study is to consider alternative information systems solutions, evaluate their feasibility, and propose the alternative most suitable to the organization. The feasibility of a proposed solution is evaluated in terms of its components.

- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 WORKING

INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

OBJECTIVES

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

2.It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed

in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow

OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.

2. Select methods for presenting information.

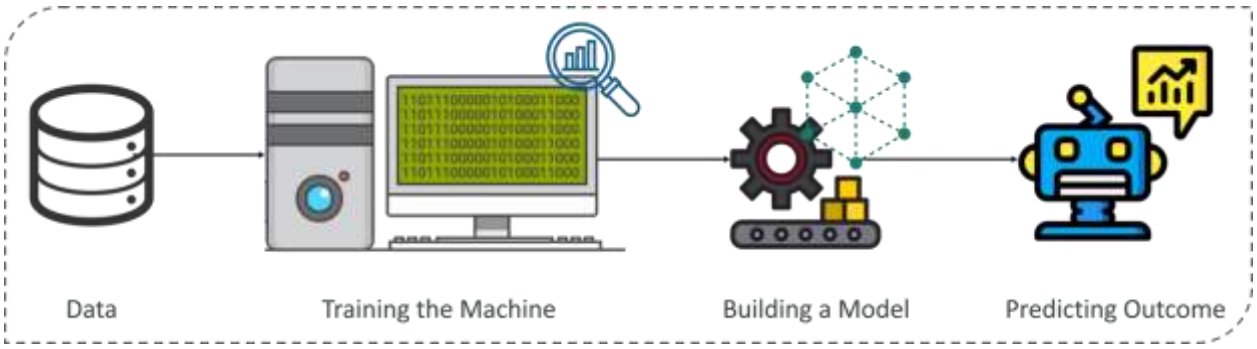
3. Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

- Convey information about past activities, current status or projections .

- Future.
- Signal important events, opportunities, problems, or warnings.
- Trigger an action.
- Confirm an action.

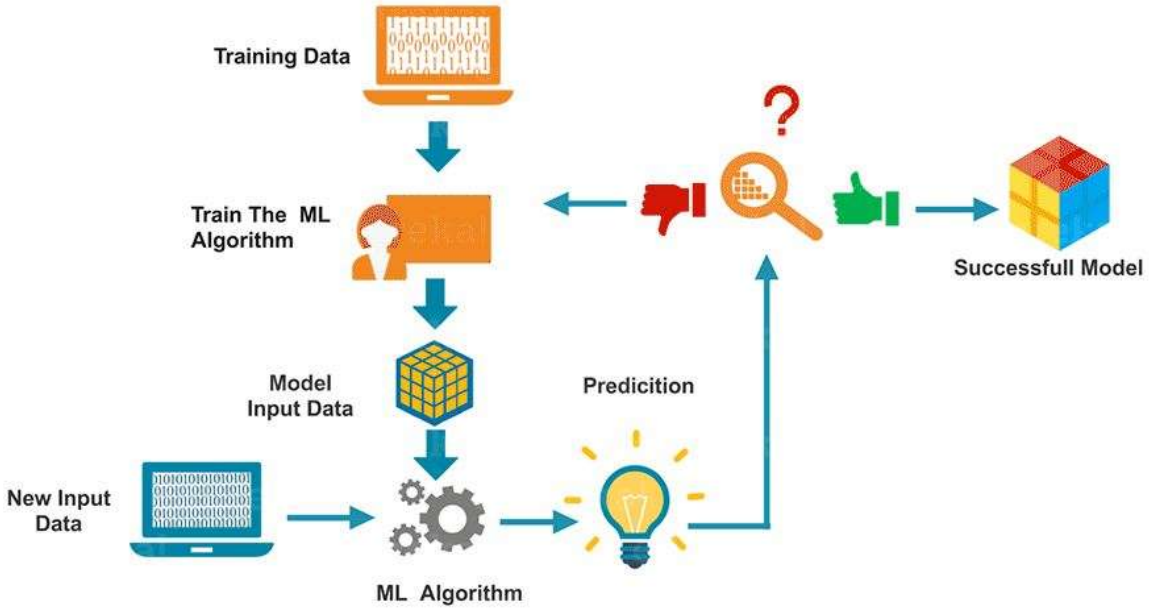
Machine Learning Process



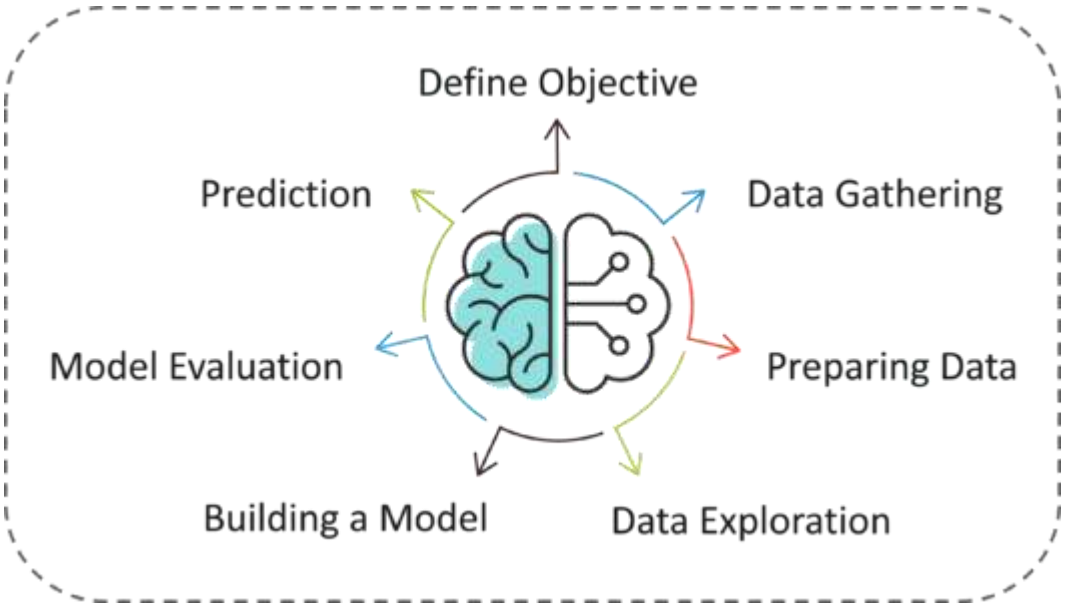
How does Machine Learning Work?

Machine Learning algorithm is trained using a training data set to create a model. When new input data is introduced to the ML algorithm, it makes a prediction on the basis of the model.

The prediction is evaluated for accuracy and if the accuracy is acceptable, the Machine Learning algorithm is deployed. If the accuracy is not acceptable, the Machine Learning algorithm is trained again and again with an augmented training data set.



The Machine Learning process involves building a Predictive model that can be used to find a solution for a Problem Statement. To understand the Machine Learning process let's assume that you have been given a problem that needs to be solved by using Machine Learning.



The below steps are followed in a Machine Learning process:

Step 1: Define the objective of the Problem Statement

At this step, we must understand what exactly needs to be predicted. In our case, the objective is to predict the possibility of rain by studying weather conditions. At this stage, it is also essential to take mental notes on what kind of data can be used to solve this problem or the type of approach you must follow to get to the solution.

Step 2: Data Gathering

At this stage, you must be asking questions such as,

- What kind of data is needed to solve this problem?
- Is the data available?
- How can I get the data?

Once you know the types of data that is required, you must understand how you can derive this data. Data collection can be done manually or by web scraping. However, if you're a beginner and you're just looking to learn Machine Learning you don't have to worry about getting the data. There are 1000s of data resources on the web, you can just download the data set and get going.

Coming back to the problem at hand, the data needed for weather forecasting includes measures such as humidity level, temperature, pressure, locality, whether or not you live in a hill station, etc. Such data must be collected and stored for analysis.

Step 3: Data Preparation

The data you collected is almost never in the right format. You will encounter a lot of inconsistencies in the data set such as missing values, redundant variables, duplicate values, etc. Removing such inconsistencies is very essential because they might lead to wrongful computations and predictions. Therefore, at this stage, you scan the data set for any inconsistencies and you fix them then and there.

Step 4: Exploratory Data Analysis

Grab your detective glasses because this stage is all about diving deep into data and finding all the hidden data mysteries. EDA or Exploratory Data Analysis is the brainstorming stage of Machine Learning. Data Exploration involves understanding the patterns and trends in the data. At this stage, all the useful insights are drawn and correlations between the variables are understood.

For example, in the case of predicting rainfall, we know that there is a strong possibility of rain if the temperature has fallen low. Such correlations must be understood and mapped at this stage.

Step 5: Building a Machine Learning Model

All the insights and patterns derived during Data Exploration are used to build the Machine Learning Model. This stage always begins by splitting the data set into two parts, training data, and testing data. The training data will be used to build and analyze the model. The logic of the model is based on the Machine Learning Algorithm that is being implemented.

Choosing the right algorithm depends on the type of problem you're trying to solve, the data set and the level of complexity of the problem. In the upcoming sections, we will discuss the different types of problems that can be solved by using Machine Learning.

Step 6: Model Evaluation & Optimization

After building a model by using the training data set, it is finally time to put the model to a test. The testing data set is used to check the efficiency of the model and how accurately it can predict the outcome. Once the accuracy is calculated, any further improvements in the model can be implemented at this stage. Methods like parameter tuning and cross-validation can be used to improve the performance of the model.

Step 7: Predictions

Once the model is evaluated and improved, it is finally used to make predictions. The final output can be a Categorical variable (eg. True or False) or it can be a Continuous Quantity (eg. the predicted value of a stock).

In our case, for predicting the occurrence of rainfall, the output will be a categorical variable.

Product Perspective

- The project mainly focuses on the basic types of crops and their nature of yield in various types of seasons.
- This project is included with a huge amount of data sets wherein almost all types of crops and their wide range of behavior and yield in various types of seasons is provided.
- These large amount of data sets helps in predicting the right crop with high amount of productivity.
- Through this, a person who does farming as their major occupation will get to know which type of plant serves them with best results.

Libraries:

Pandas:

Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. The name Pandas is derived from the word Panel Data – an Econometrics from Multidimensional data.

In 2008, developer Wes McKinney started developing pandas when in need of high performance, flexible tool for analysis of data.

Prior to Pandas, Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data — load, prepare, manipulate, model, and analyze.

Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

Key Features of Pandas

- Fast and efficient DataFrame object with default and customized indexing.
- Tools for loading data into in-memory data objects from different file formats.
- Data alignment and integrated handling of missing data.
- Reshaping and pivoting of date sets.
- Label-based slicing, indexing and subsetting of large data sets.
- Columns from a data structure can be deleted or inserted.
- Group by data for aggregation and transformations.
- High performance merging and joining of data.
- Time Series functionality.

Numpy:

NumPy is a Python package. It stands for 'Numerical Python'. It is a library consisting of multidimensional array objects and a collection of routines for processing of array.

Numeric, the ancestor of NumPy, was developed by Jim Hugunin. Another package Numarray was also developed, having some additional functionality. In 2005, Travis Oliphant created NumPy package by incorporating the features of Numarray into Numeric package. There are many contributors to this open source project.

Operations using NumPy

- Using NumPy, a developer can perform the following operations –
- Mathematical and logical operations on arrays.
- Fourier transforms and routines for shape manipulation.
- Operations related to linear algebra. NumPy has in-built functions for linear algebra and random number generation.

Keras:

Deep learning is one of the major subfield of machine learning framework. Machine learning is the study of design of algorithms, inspired from the model of human brain. Deep learning is becoming more popular in data science fields like robotics, artificial intelligence (AI), audio & video recognition and image recognition. Artificial neural network is the core of deep learning methodologies. Deep learning is supported by various libraries such as Theano, TensorFlow, Caffe, Mxnet etc., Keras is one of the most powerful and easy to use python library, which is built on top of

popular deep learning libraries like TensorFlow, Theano, etc., for creating deep learning models.

DJANGO:

What is Django?

Django is a Python-based web framework which allows you to quickly create web application without all of the installation or dependency problems that you normally will find with other frameworks.

When you're building a website, you always need a similar set of components: a way to handle user authentication (signing up, signing in, signing out), a management panel for your website, forms, a way to upload files, etc. Django gives you ready-made components to use.

Why Django?

It's very easy to switch database in Django framework.

It has built-in admin interface which makes easy to work with it.

Django is fully functional framework that requires nothing else.

It has thousands of additional packages available.

It is very scalable.

Popularity of Django

Django is used in many popular sites like as: Disqus, Instagram, Knight Foundation, MacArthur Foundation, Mozilla, National Geographic etc. There are more than 5k online sites based on the Django framework. (Source)

Sites like Hot Frameworks assess the popularity of a framework by counting the number of GitHub projects and StackOverflow questions for each platform, here Django is in 6th position. Web frameworks often refer to themselves as "opinionated" or "un-opinionated" based on opinions about the right way to handle any particular task. Django is somewhat opinionated, hence delivers the in both worlds(opinionated & un-opinionated).

Features of Django

Versatility of Django

Django can build almost any type of website. It can also work with any client-side framework and can deliver content in any format such as HTML, JSON, XML etc. Some sites which can be built using Django are wikis, social networks, new sites etc

Security

Since Django framework is made for making web development easy, it has been engineered in such a way that it automatically do the right things to protect the website. For example, In the Django framework instead of putting a password in cookies, the hashed password is stored in it so that it can't be fetched easily by hackers.

Scalability

Django web nodes have no stored state, they scale horizontally – just fire up more of then when you need them. Being able to do this is the essence of good scalability. Instagram and Disqus are two Django based products that have millions of active users, this is taken as an example of the scalability of Django.

Portability

All the codes of the Django framework are written in Python, which runs on many platforms. Which leads to run Django too in many platforms such as Linux, Windows and Mac OS.

Tomcat Server:

Introduction

TomCat, in its simplest concept, is a web server. Well, it's more than that. While it can serve static files, its primary purpose is to act as a 'servlet container' that serves Java web applications. It can process [.jsp] files, which are like PHP scripts but for Java, and it can also run [Java Servlets], which are classes that process the GET, POST, and other HTTP requests. TomCat will listen on TCP ports for HTTP requests, and route requests to your Java classes, JSP files, or static files. It is also possible to embed TomCat in to a standalone application, but that is out of the scope of this document.

Beginner level Java knowledge is expected, although there is virtually no Java code in this tutorial specfically. This will cover basic installation, configuration, and admin aspects. This is aimed at beginner to intermediate developers who want to learn about basic Java web app hosting. You should either be a system administrator setting up TomCat or a developer learning how to set up TomCat.

I would always recommend using the official documentation at http://tomcat.apache.org/ for the latest and most accurate information. This is meant as a quick start guide that includes the things I think are important to know to get started. There are a number of features that are

not covered here. For example, TomCat can be embedded inside your application.One popular alternative to TomCat is Eclipse Jetty. It is often used when you need a lightweight embedded servlet container.

Deploying Django with Apache and mod_wsgi is a tried and tested way to get Django into production.

mod_wsgi is an Apache module which can host any Python WSGI application, including Django. Django will work with any version of Apache which supports mod_wsgi.

The official mod_wsgi documentation is your source for all the details about how to use mod_wsgi. You'll probably want to start with the installation and configuration documentation.

About SQLite

SQLite is an in-process library that implements a self-contained, serverless, zero-configuration, transactional SQL database engine. The code for SQLite is in the public domain and is thus free for use for any purpose, commercial or private. SQLite is the most widely deployed database in the world with more applications than we can count, including several high-profile projects.

SQLite is an embedded SQL database engine. Unlike most other SQL databases, SQLite does not have a separate server process. SQLite reads

and writes directly to ordinary disk files. A complete SQL database with multiple tables, indices, triggers, and views, is contained in a single disk file. The database file format is cross-platform - you can freely copy a database between 32-bit and 64-bit systems or between big-endian and little-endian architectures. These features make SQLite a popular choice as an Application File Format. SQLite database files are a recommended storage format by the US Library of Congress. Think of SQLite not as a replacement for Oracle but as a replacement for fopen()

SQLite is a compact library. With all features enabled, the library size can be less than 500KiB, depending on the target platform and compiler optimization settings. (64-bit code is larger. And some compiler optimizations such as aggressive function inlining and loop unrolling can cause the object code to be much larger.) There is a tradeoff between memory usage and speed. SQLite generally runs faster the more memory you give it. Nevertheless, performance is usually quite good even in low-memory environments. Depending on how it is used, SQLite can be faster than direct filesystem I/O.

SQLite is very carefully tested prior to every release and has a reputation for being very reliable. Most of the SQLite source code is devoted purely to testing and verification. An automated test suite runs millions and millions of test cases involving hundreds of millions of individual SQL statements and achieves 100% branch test coverage. SQLite responds gracefully to memory allocation failures and disk I/O errors. Transactions are ACID even if interrupted by system crashes or power failures. All of this is verified by the automated tests using special test harnesses which simulate system failures. Of course, even with all this testing, there are still bugs. But unlike some similar projects (especially commercial

competitors) SQLite is open and honest about all bugs and provides bugs lists and minute-by-minute chronologies of code changes.

The SQLite code base is supported by an international team of developers who work on SQLite full-time. The developers continue to expand the capabilities of SQLite and enhance its reliability and performance while maintaining backwards compatibility with the published interface spec, SQL syntax, and database file format. The source code is absolutely free to anybody who wants it, but professional support is also available.

The SQLite project was started on 2000-05-09. The future is always hard to predict, but the intent of the developers is to support SQLite through the year 2050. Design decisions are made with that objective in mind.

We the developers hope that you find SQLite useful and we entreat you to use it well: to make good and beautiful products that are fast, reliable, and simple to use. Seek forgiveness for yourself as you forgive others. And just as you have received SQLite for free, so also freely give, paying the debt forwar

CHAPTER 5

5.1 Conclusion

With this model, we have created a rudimentary model that is able to forecast to a certain extent. Even though the model is not perfect, we have one that can approximate to the past data pretty well. But for new data, we would require more parameter tuning.

There exists great potential to improve sales forecasting accuracy in the Ecommerce domain. One good opportunity is to utilize the correlated and similar sales patterns available in a product portfolio. In this paper, we have introduced a novel demand forecasting framework based on LSTMs that exploits non-linear relationships that exist in the E-commerce business. We have used the proposed approach to forecast the sales demand by training a global model across the items available in a product assortment hierarchy. Our developments also present several systematic grouping strategies to our base model, which are in particular useful in situations where product sales are sparse. Our methodology has been evaluated on a real-world E-commerce database from Walmart.com. To demonstrate the robustness of our framework, we have evaluated our methods on both category level and super-department level datasets. The results have shown that our methods have outperformed the state-of-the-art univariate forecasting techniques. Furthermore, the results indicate that E-commerce product hierarchies contain various cross-product demand patterns and correlations are available, and approaches to exploit this information are necessary to improve the sales forecasting accuracy in this domain.

So, we have discussed four different steps in data cleaning to make the data more reliable and to produce good results. After properly completing the Data Cleaning steps, we'll have a robust dataset that avoids many of the most common pitfalls. This step should not be rushed as it proves very beneficial in the further process.

Feature Scaling is a technique to standardize the independent features present in the data in a fixed range. It is performed during the data pre-processing.

APPENDICES

A. **SOURCE CODE**

Design Template Code:

```
<div class="contact-form">
	<h3 class="style">Fill The Following Details</h3>
<h3 class="style"><font color="red">{{ output }}</font></h3>
<h3 class="style"><font color="red">{{ error }}</font></h3>
	<form method="post" action="/predictaction/">
		{% csrf_token %}
<div>
			<span><label>SELECT DISTRICT</label></span>
			<span>
				<select name="District_Name">
					{% for key, value in districts.items %}
			<option value="{{ value }}" class="textbox">{{ key
}}</option>
			 {% endfor %}
				</select>
			</span>
		</div>
<div>
			<span><label>SELECT SEASON</label></span>
			<span>
```

```
                        <select name="Season">
                        <option value="1"
                        class="textbox">Winter</option>

                <option value="2" class="textbox">Summer</option>

<option value="3" class="textbox">Rainy</option>

                        </select>

                </span>

        </div>

        <div>

                <span><label>LAND AREA</label></span>

                <span><input name="Area" type="text"
        class="textbox"></span>

        </div>

        <div>

                <span><label>Temperature</label></span>

                <span><input name="temperature" type="text"
        class="textbox"></span>

        </div>

        <div>

                <span><label>Humidity</label></span>

                <span><input type="text"
        name="humidity"></input></span>

        </div>

<div>

                <span><label>PH Value</label></span>

                <span><input name="ph" type="text"
        class="textbox"></span>

        </div>

        <div>

                <span><label>Rain Fall</label></span>
```

```
                <span><input name="rainfall" type="text"
            class="textbox"></span>
            </div>
            <div>
                <span><input type="submit" value="Predict"></span>
            </div>
        </form>
</div>
```

Django Code:

```
import pandas as pd
from django.shortcuts import render
from yieldforecasting.appconstants import path1,path2
from yieldforecasting.forms import DataForm
from yieldforecasting.service import forecast, loaddata
def predict(request):
    data=loaddata()
    return render(request, 'predict.html', {"districts":data})
def predictAction(request):
    if request.method == "POST":
        dataForm = DataForm(request.POST)
        if dataForm.is_valid():
            District_Name = dataForm.cleaned_data["District_Name"]
            Season = dataForm.cleaned_data["Season"]
            Area = dataForm.cleaned_data["Area"]
            temperature = dataForm.cleaned_data["temperature"]
            humidity = dataForm.cleaned_data["humidity"]
            ph = dataForm.cleaned_data["ph"]
            rainfall = dataForm.cleaned_data["rainfall"]
```

```
result=forecast(District_Name,Season,Area,temperature,humidity,ph,rainfall)
            output="Suggested Crop:"+str(result[0])+" and Predicted Yield:"+str(result[1])
            data = loaddata()
            return render(request, 'predict.html', {"output":output,"districts": data})
        else:
          return render(request, 'predict.html', {"error": "Invalid Data"})
    else:
        return render(request, 'predict.html', {"error": "Invalid Request"})
```

Machine Learning Prediction Code:

```
import pandas as pd

from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn import linear_model

import numpy as np

from yieldforecasting.appconstants import path1,path2

def loaddata():

    df2 = pd.read_csv(path2)
    #
==========================================================
    districts = df2['District_Name']
```

```
    districts = districts.drop_duplicates()
    district_dict = dict()
    #
==========================================================
========================

    i = 1
    for dist in districts:
        district_dict.update({dist: i})
        i = i + 1

    print(district_dict)

    return district_dict

def
forecast(District_Name,Season,Area,temperature,humidity,ph,rainfall):

    df1 = pd.read_csv(path1)
    df2 = pd.read_csv(path2)
    #
==========================================================
====================

    crops1 = df1['Crop']
    crops2 = df2['Crop']
    districts = df2['District_Name']

    crops1 = crops1.drop_duplicates()
    crops2 = crops2.drop_duplicates()
```

```
    districts = districts.drop_duplicates()


    crop1_dict = dict()
    crop2_dict = dict()
    district_dict = dict()


    #
==========================================================
====================
    i = 1
    for crop in crops1:
        crop1_dict.update({crop: i})
        i = i + 1


    df1['Crop'] = df1['Crop'].map(crop1_dict)


    #
==========================================================
========================


    i = 1
    for crop in crops2:
        crop2_dict.update({crop: i})
        i = i + 1


    df2['Crop'] = df2['Crop'].map(crop2_dict)


    #
==========================================================
========================
```

```
    i = 1
    for dist in districts:
        district_dict.update({dist: i})
        i = i + 1

    df2['District_Name'] = df2['District_Name'].map(district_dict)

    #
============================================================
==========================

    df2['Crop'] = df1['Crop']

    #
============================================================
==========================

    feature_cols = ["temperature", "humidity", "ph", "rainfall"]
    X = df1[feature_cols]  # Features
    y = df1.Crop  # Target variable

    # Split X and y into train and test sets
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=1)

    clf = DecisionTreeClassifier()
    clf = clf.fit(X_train, y_train)

    # y_pred = clf.predict(X_test)
```

```
    my_pred = clf.predict([[temperature, humidity, ph, rainfall]])
    predicted_crop = my_pred[0]

    def get_key(val):
        for key, value in crop1_dict.items():
            if val == value:
                return key

    finacrop=get_key(predicted_crop)
    #
==========================================================
=======================

    df2[:] = np.nan_to_num(df2)

    feature_cols = ["District_Name", "Season", "Crop", "Area"]

    X = df2[feature_cols]  # Features
    y = df2.Production  # Target variable

    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=1)

    regr = linear_model.LinearRegression()
    regr.fit(X_train, y_train)

print(type(District_Name),type(Season),type(predicted_crop),type(Area))
```

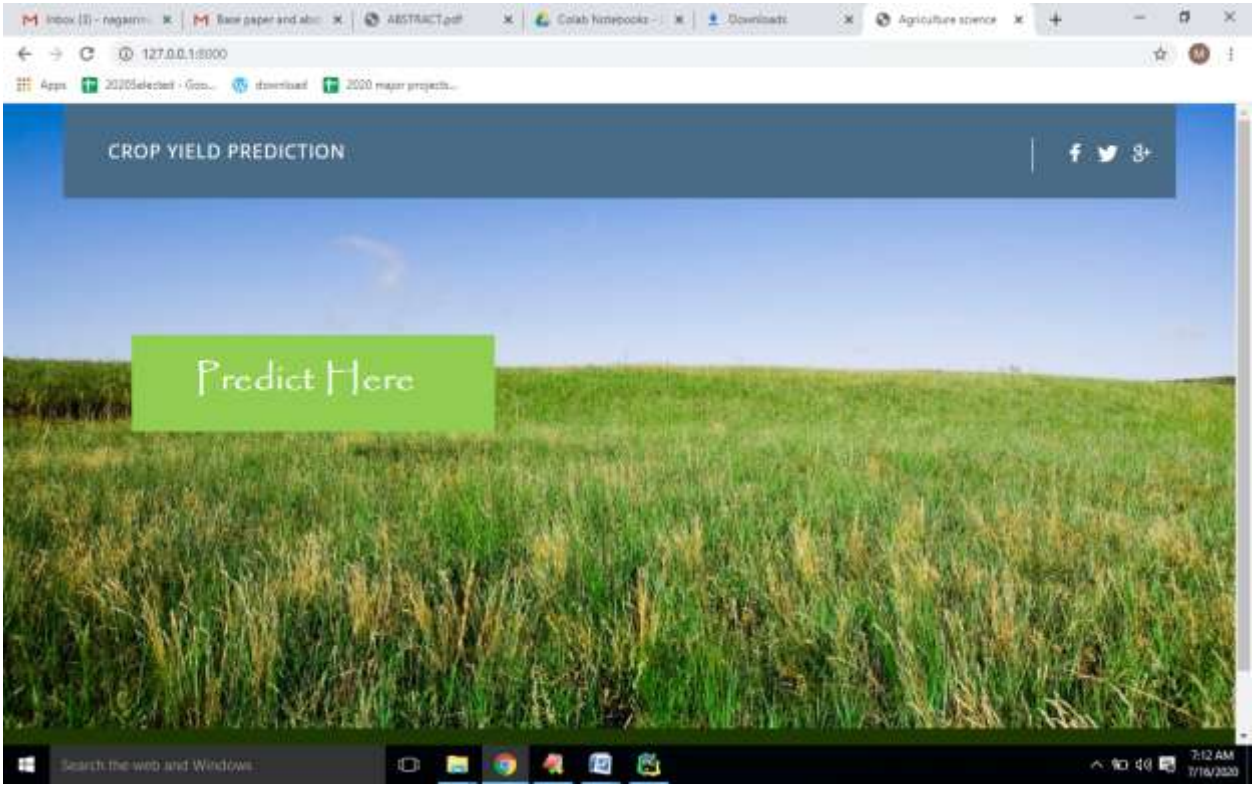
```
    y_pred = regr.predict([[District_Name, Season, predicted_crop,
Area]])


    return [finacrop,y_pred[0]]
```

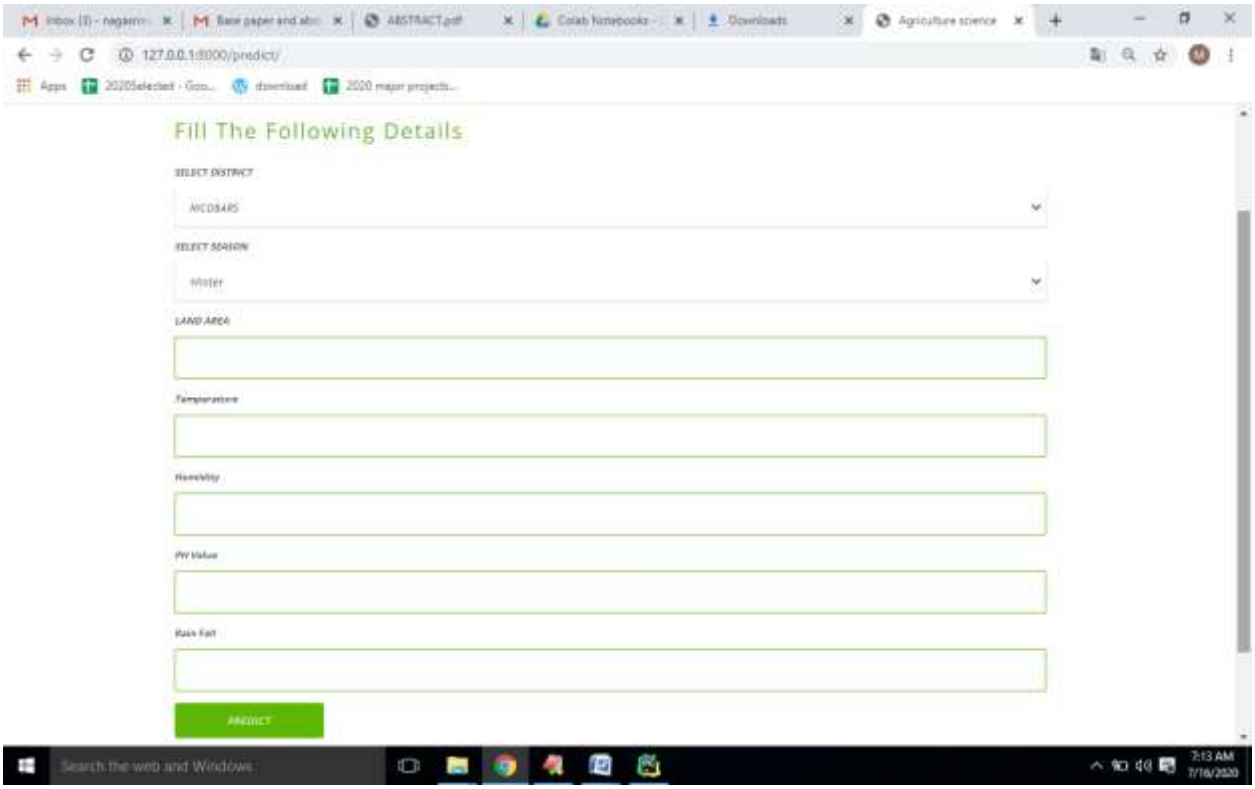
ULRS

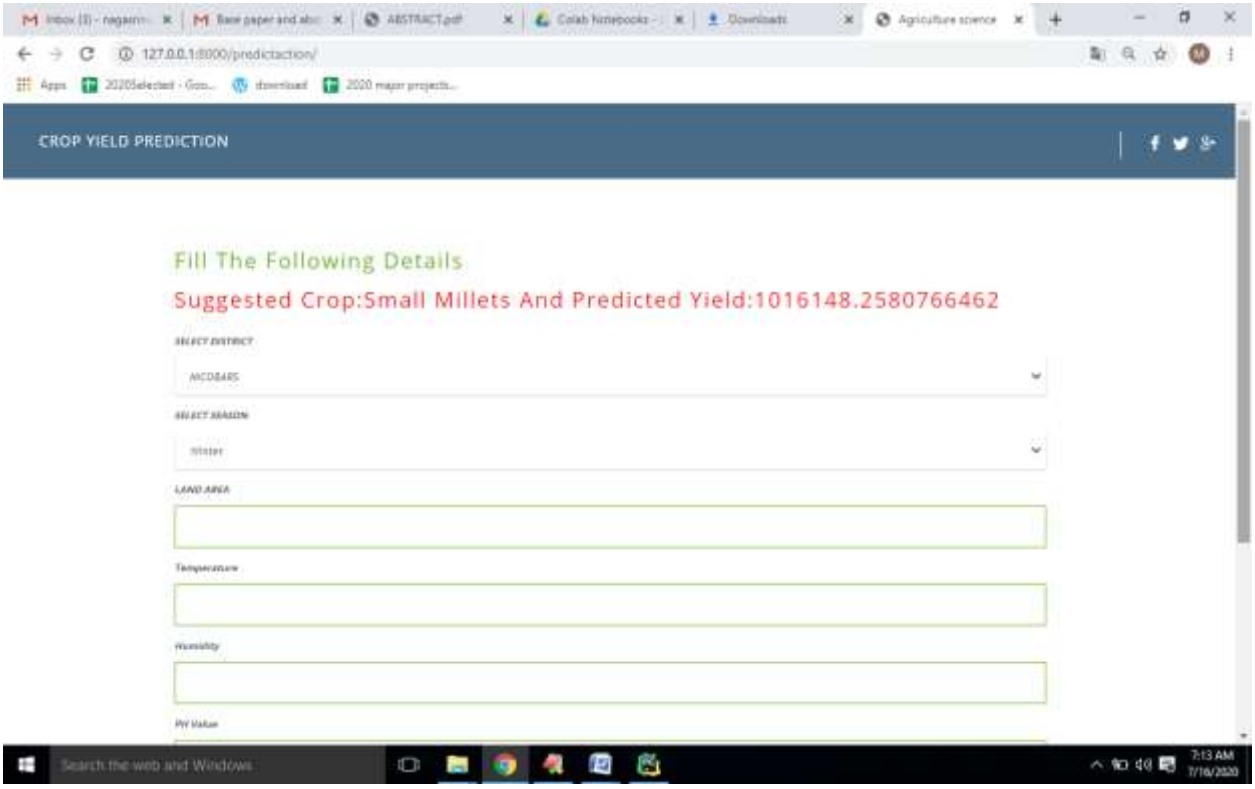
```
path('',TemplateView.as_view(template_name =
'index.html'),name='index'),

path('index/',TemplateView.as_view(template_name =
'index.html'),name='index'),

path('predict/',predict,name='predict'),

path('predictaction/',predictAction,name='predictaction'),
```

B. SCREENSHOTS

Original Data Set




Fill The Following Details
SELECT DISTRICT
ANDAMANS
SELECT SEASON
Winter
LAND AREA
Temperature
Humidity
PH Value
Rain Fall
PREDICT


CROP YIELD PREDICTION
Fill The Following Details
Suggested Crop:Small Millets And Predicted Yield:1016148.2580766462
SELECT DISTRICT
ANDAMANS
SELECT SEASON
Winter
LAND AREA
Temperature
Humidity
PH Value

Test cases

Tested	Test name	Inputs	Expected output	Actual Output	Result
1	load dataset	dataset	dataset loaded	successfully loaded	pass

Tested	Test name	Inputs	Expected output	Actual Output	Result
2	Splitting dataset into training and validation set	dataset	spitted to training and validation	successfully spitted to train data and validation data	pass

Tested	Test name	Inputs	Expected output	Actual Output	Result
4	Model creation	Keras module with lstm	model created	successfully model created	pass

Tested	Test name	Inputs	Expected output	Actual Output	Result
5	training set evaluation	train data	training done	successfully trained	pass

Tested	Test name	Inputs	Expected output	Actual Output	Result

6	Validating test dataset	test data	test dataset validated	successfully validated	pass

Tested	**Test name**	**Inputs**	**Expected output**	**Actual Output**	**Result**
7	getting accuracy	test dataset	accuracy in percentage	successfully got accuracy	pass

Tested	**Test name**	**Inputs**	**Expected output**	**Actual Output**	**Result**
8	load test set	test data	detecting price	successfully detected	pass

Tested	**Test name**	**Inputs**	**Expected output**	**Actual Output**	**Result**
9	print result	test results	printing test results or save in csv file	Successfully saved	success

References

1. Hyndman, R. et al., 2008. Forecasting with Exponential Smoothing: The State Space Approach, Springer Science & Business Media.

2. Box, G.E.P. et al., 2015. Time Series Analysis: Forecasting and Control, John Wiley & Sons.

3. Box, G.E.P. & Cox, D.R., 1964. An Analysis of Transformations. Journal of the Royal Statistical Society. Series B, Statistical methodology, 26(2), pp.211252.

4. Yeo, J. et al., 2016. Browsing2purchase: Online Customer Model for Sales Forecasting in an E-Commerce Site. 25th Int.Conf.Comp. on World Wide Web.

5. Ramanathan, U., 2013. Supply chain collaboration for improved forecast accuracy of promotional sales. Int. Journal of Operations & Production Management.

6. Kulkarni, G., Kannan, P.K. & Moe, W., 2012. Using online search data to forecast new product sales. Decision support systems, 52(3), pp.604611.

7. Zhao, K. & Wang, C., 2017. Sales Forecast in E-commerce using Convolutional Neural Network. arXiv [cs.LG].

8. Seeger, M.W., Salinas, D. & Flunkert, V., 2016. Bayesian Intermittent Demand Forecasting for Large Inventories. In Proceedings of the 29th NIPS.

9. Snyder, R., Ord, J.K. & Beaumont, A., 2012. Forecasting the intermittent demand for slow-moving inventories: A modelling approach. IJF, 28(2), pp.485496.

10. Zhang, G., Patuwo, B.E. & Hu, M.Y., 1998. Forecasting with artificial neural networks: The state of the art. International journal of forecasting, 14(1), pp.3562.

11. Yan, W., 2012. Toward automatic time-series forecasting using neural networks. IEEE transactions on neural networks and learning systems, 23(7), pp.10281039.

12. Zimmermann, H.-G., Tietz, C. & Grothmann, R., 2012. Forecasting with RNNs: 12 Tricks. In Neural Networks. Lecture Notes in Computer Science. pp. 687707.

13. Trapero, J.R., Kourentzes, N. & Fildes, R., 2015. On the identification of sales forecasting models in the presence of promotions. The Journal of the ORS.

14. Borovykh, A., Bohte, S. & Oosterlee, C.W., 2017. Conditional Time Series Forecasting with Convolutional Neural Networks arXiv [cs.AI].

15. Flunkert, V., Salinas, D. & Gasthaus, J., 2017. DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks. arXiv [cs.AI].

16. Wen, R. et al., 2017. A Multi-Horizon Quantile Recurrent Forecaster. arXiv [stat.ML].

C. PLAGIARISM REPORT

CROP YIELD PREDICTION & RECOMM...

ORIGINALITY REPORT

9%	6%	5%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.mdpi.com Internet Source	2%
2	Submitted to Rajiv Gandhi Proudyogiki Vishwavidyalaya Student Paper	1%
3	proceedings.mlr.press Internet Source	1%
4	toc.proceedings.com Internet Source	1%
5	B. Vinoth, N. M. Elango. "An effective data mining techniques based optimal paddy yield cultivation: a rational approach", Paddy and Water Environment, 2021 Publication	1%
6	ijarcce.com Internet Source	1%
7	"Table of Contents", 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021 Publication	1%

D. JOURNAL PAPER

A MACHINE LEARNING APPROACH For CROP YIELD PREDICTION & RECOMMENDATION

Budha Pretesh [1],Ch Sai Teja [2],Murari Devakannan Kamalesh [3] UG

Student[1][2], Department of Computer Science,

Sathyabama Institute of Science and Technology,Chennai,600119,India

Professor[3],Department of Computer Science,

Sathyabama Institute of Science and Technology,Chennai,600119,India

ABSTRACT

Fertilizer value updating has a positive practical significance for guiding agricultural production and for notifying the change in market rate of fertilizer for farmer. This task utilizes information examination innovation to change costs by warning. An immediate notice will be sent when how much compost is moved to the market. Different soils. Hence, the nature of compost is dictated by the interaction. Essentially, inferior quality, excellent modern composts are delivered. Utilizing a successive majority permits you to settle on better choices about theories by utilizing numerous classifications. Moreover, the positioning system is utilized to settle on choices about the choice of results. The framework is utilized to uncover the cost of manure to further develop compost.

Keywords— K-Nearest Neighbour (KNN), Naïve bayes, SVM Algorithm, Crop yield prediction

INTRODUCTION:

The accomplishment of data innovation on the planet will assist ranchers with recognizing and foster their latent capacity. Data trade is significant data sharing and on account of ranchers, both in the conventional way. The craving to share data mirrors the open disposition of ranchers. This open technique decides the level and size of data trade. We assemble sites utilizing web advances like html and css, gather information from numerous sources, put it down, foresee compost costs, and do disconnected exploration. Bundle and rundown are given in the compost list. Send the necessary data and offer it with the ranchers who gather and store the data on the MySQL server. We are a product that consequently sends state-of-the-art farming data. This disposes of the requirement for ranchers to make a trip nearer to urban areas and towns to get modern data. We will have an AI calculation at anticipating manure costs in the following two months. With that in mind, we use vector machine (SVM), Naïve Baye (NB) and K-Nearest Neighbour (KNN) calculations to decide the expense of manure for ranchers to utilize. Also, the arranging system is utilized to settle on dynamic choices.

EXISTING SYSTEM:

In an existing system either only crop recommendation or crop yield prediction is only present. Some of the crop recommendation systems in the market are with very low accuracy

Many existing crop yield prediction are based on only area and do not concentrate on soil type and temperature. The main factor in predicting crop yield is the type of crop which many crop yield predictions do not concentrate.

A. DISADVANTAGES OF EXISTING SYSTEM:

- Efficiency is low.
- Repetition of work.

LITERATURE SURVEY:

[1]Title: **Crop Yield Prediction using Machine Learning Algorithm**

D.Jayanarayana Reddy; M. Rudra Kumar

Agribusiness is the foundation of India's economy, with in excess of 50% of the populace occupied with cultivating. Environmental change, environmental change and other ecological variables altogether affect farming wellbeing. Machine learning (ML) assumes a significant part as it is an apparatus for Crop Yield Prediction (CYP) independent direction.

[2]Title: **Crop Yield Prediction in Precision Agriculture**

Prof. Dr. Miklós Neményi

The prediction of crop yield is based on soil, meteorological, environmental, and crop parameters. Decision support models are broadly used to extract significant crop features for prediction. Precision agriculture focuses on monitoring (sensing technologies), management systems, variable rate technologies, and responses to inter- and intervisibility cropping systems.

[3]Title: **Crop Recommender System Using Machine Learning Approach**

Shilpa mangesh pande; prem kumar ramesh; anmol; b. R aishwarya; karuna rohilla

There is no question that farming and enterprises are connected to the jobs of country Indians. This is one of the principle justifications for why peripheral ranchers in India end it all

[4]Title: **AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms**

ZeelDoshi; Subhash Nadkarni; Rashi Agrawal

horticulture is a significant commitment to the Indian economy. The huge number of individuals living in India relies upon how they live in horticulture. Numerous Indian ranchers accept that they can pick plants to plant at a given time.

[5]Title: A Review on Data Mining Techniques for Fertilizer Recommendation, 2018

Authors: Jignasha M. Jethva, Nikhil Gondaliya, Vinita Shah

At the point when the dirt is insufficient in supplements, add compost to decrease it. A typical issue in farming is excrement determination and fertilizer expansion. Extreme development or absence of manure can harm vegetation and diminish efficiency. This record sums up the different techniques for removing information used to develop a bunch of modern manure soils.

[6]Title: A Survey on Data Mining Techniques in Agriculture, 2015

Author: M.C.S.Geetha

Horticulture is important for the economies of emerging nations, particularly India. Mining assumes a significant part in decision-production in numerous spaces of farming. It inspects the job of data mining in the farming area and work corresponding to a couple of creators in the rural area. It likewise checks out various methods of acquiring data to resolve numerous agrarian issues. This paper integrates the work of several authors in a single place so it is valuable for specialists to get data of current situation of data mining systems and applications in context to farming field.

[7]Title: AgroNutri Android Application, 2016

Authors: S. Srija, R. Geetha Chanda, S.Lavanya, Dr. M. Kalpana Ph.D

This paper communicates the idea regarding the making of AgroNutri an android application that helps in conveying the harvest particular fertilizer amount to be applied. The future scope of the AgroNutri is that GPRS can be included so that according to location nutrients are suggested. Further this application would be incorporated as a piece of the accuracy agriculture wherein sensors can be utilized to discover the measure of NPK present in the dirt and that sum can be deducted from the suggestion and giving us the exact measure of supplements to be added.

[8]Title: Machine Learning: Applications in Indian Agriculture, 2016

Authors: Karandeep Kaur

Agribusiness was an area that was contrary with innovation and its turn of events. Indian ranchers should comply with the common principles. AI is the fundamental thought utilized in all data sources and results. He has utilized his capacity in basic science and programming. Mechanical preparing figures have incredibly worked on the craft of AI

and incorporate sensor-based parts utilized in coordinated agribusiness. This paper analyses the different employments of AI in horticultural fields. It gives experiences into the issues looked by ranchers and how to settle them utilizing this technique.

PROPOSED SYSTEM:

Crop Recommendation System

If farmers are confused of what crop to be cultivated they can open this application and ask for a recommendation. For getting crop recommended they need to give inputs such as nitrogen, phosphorus, temperature, PH, humidity and rainfall.

Based on these inputs a suitable crop will be recommended by the application

Crop yield prediction

Farmers can predict the amount of crop by various factors like area, type of crop, temperature, rainfall, and soil type. This prediction is done using regression algorithm.

By this prediction farmers can calculate the amount of profit they will getting. This helps farmers in their financial management.

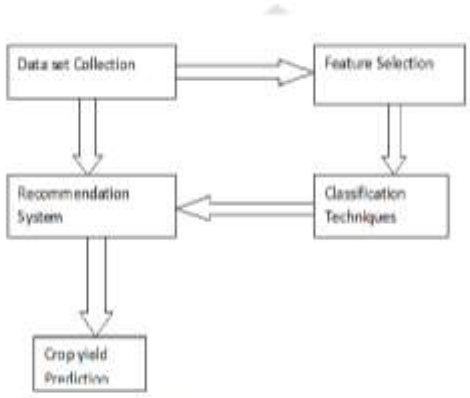


Figure 1: Proposed Architecture for Crop Recommendation

A. ADVANTAGES OF PROPOSED SYSTEM:

- Useful for individuals a long way from urban communities and towns.
- Incredible time.
- Limit checking.

ALGORITHM USED

1. K-Nearest Neighbour (KNN) Algorithm:

K, N, as associate degree example is quite a foundation for teaching or teaching a lazy man concerning grace: it's a lot of concerning obtaining getting ready to wherever the calculation is postponed, and performance of the partition. K, N, machine learning algorithmic programs of the algorithm could be a terribly straightforward factor. and also the proximity arising from the category provide (class of K n) to the worth of the article (to proceed K N) is verified.

STEP 1: BEGIN

STEP 2: Input: D =

STEP 3: another instance of arranging x = (x1 ... Xn)

STEP 4: Count (xi, ci) d (xi, x) for each case composed.

Stage 5: Separate d (xi, x) from base to top, (I = 1 .... N)

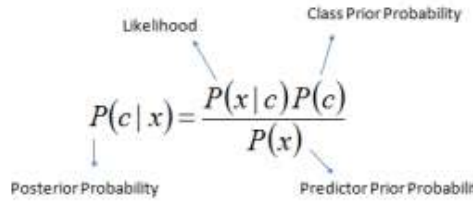
STEP 6: x: Select K for instance close to Dkx

STEP 7: Scores x Dkx general classification

STEP 8: Completion

1. *Naïve bayes Algorithm:*

P(X) due to an earlier case. The technique relies on split Bayes associated with the conclusion of the first step on the assumption of free predictors. In the presence of the fixed function of defined limits I am in the presence of a simple categorizer Bayes too much foreign matter, and the other part of the bed. Even if it is the fruit of the well of the well - to shine and the properties of each other's special occasions, a companion of the opposites of one, or to confer the degree of his evil, whence it is said, `` which is good. Words for a Naive Bayes is a simple example, so that significantly terribly useful, and huge sets for the sake of knowledge. Simplicity is still attached to a more subtle kind of nice Bayes, the developer thought.



$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Above,

- P (c | x) offered the prophets the last mechanical chance (c, objective) (x, characteristic).
- P (c) is the main chance to watch out.
- P (x | c) is the capacity to anticipate the stage.

1. *SVM Algorithm:*

SVM upholds vector machines. For an informational index comprising of choices designed on an introduced name, the A SVM records models that anticipate another example order. Relegate other level/data displayed in classification 1. Assuming there are just two classifications, it tends to be shown as a paired SVM list. Here are a few kinds of SVM:

- SVM line
- Lines without SVM lines

SVM Linear Classifier:

As far as enlistment, we will more often than not accept that the mentor gives a model at home. These information focuses are planned to overcome any issues. Hyperplane forecast is straightforwardly partitioned into two phases. The main thing to do when planning a hyperplane is to diminish the separation from the hyperplane to the closest information in two stages. The hyper-plane outline is displayed as the greatest hyper-plane.

SVM Non-Linear Classifier:

Our data bundles are broadly appropriated all over the planet. Getting this data from totally various classes of hyperplants ought not be viewed as a decent choice. That is the reason Vapnik recommended making a nonlinear classifier utilizing a hyper-plane stunt. In the nonlinear SVM list, information focuses are relied upon to surpass the breaking point.

Examples of SVM boundaries:

In this section, we will figure out how to pick the best hyperplan to execute. We will show you

Category 2 data. The classes are displayed in triangles and circles.

Case 1:

- Take a gander at the issue in Figure 2 and the data in the two unique classes. Presently we need to observe a decent hyper plane that can isolate the two classifications.
- For this situation, see Figure 1. on the option to see as the proper hyper plane In SVM, we attempt to build the distance between the hyper-plane and the closest information. This is known as an edge.
- Arrangement 1 is restricted, so it is more than the distance between the left and right sides of the example. So, our most elevated hyperplan edge will be "first".

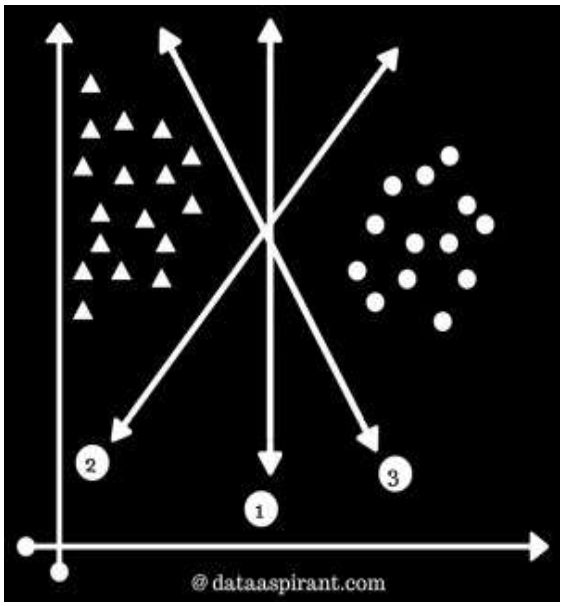


Fig.1

Case 2:

- In Figure 2, we think about two distinct classes of media. Presently we need to find a decent hyperplane that can separate between the two classes.
- Information for every class is circulated to the left or right. We will probably pick a hyperplan that can separate between classes for most extreme contrasts.
- For this situation, the choice limits are ordered, yet the limits of choice 1 demonstrate the most extreme distinction between \ bigtriangleup and \ bigcirc.

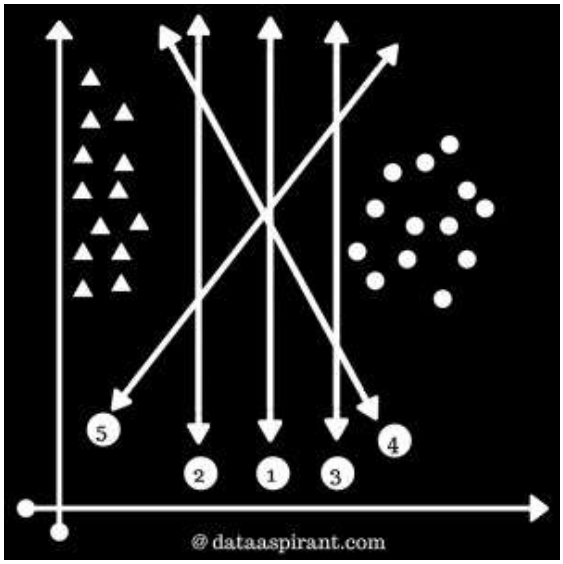
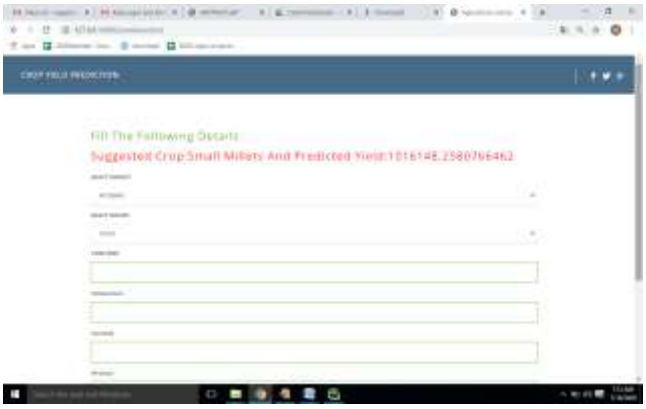


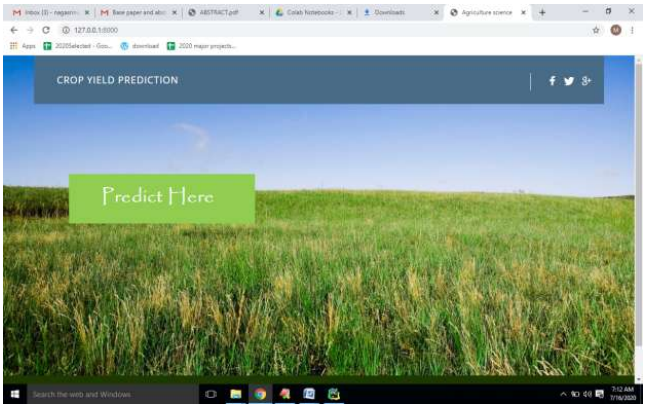
Fig.2

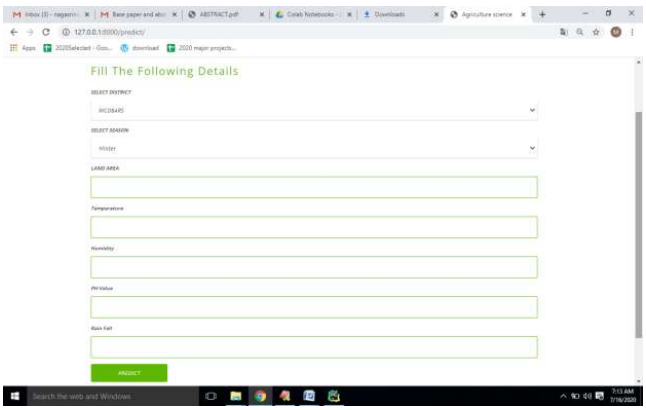
CONCLUSION &Results :

This open methodology decides the level and degree of data trade. Enormous scope logical innovation can work on the exhibition of modern composts. The venture has fostered another strategy for estimating the cost of modern manures dependent on compost costs. The thought is to utilize a majority that isolates them to foresee. Utilizing a successive majority permits you to settle on better choices about speculations by utilizing

various classifications. Also, the positioning system is utilized to settle on choices about the choice of results. The framework is utilized to uncover the cost of manure to further develop compost.







REFERENCES:

[1] Manpreet Kaur, Heena Gulati, Harish Kundra, “Data Mining in Agriculture on Crop Price Prediction: Techniques and Applications”, International Journal of Computer Applications, Volume 99– No.12, August 2014.

[2] J. Meng, “Research on the cost of agricultural products circulation and its control under the new normal economic development,” Commercial Times, no. 23, pp. 145147, 2016.

[3] A. Kaloxylos et al., “Farm management systems and the future Internet era,” Comput. Electron. Agricult., vol. 89, pp. 130–144, Nov. 2012.

[4] N. N. Li, T. S. Li, Z. S. Yu, Y. Rui, Y. Y. Miao, and Y. S. Li, “Factors influencing farmers’ adoption of new technology based on Logistic-ISM model-a case study of potato planting technology in Dingxi City, Gansu Province,” Progress in Geography, vol. 33, no. 4, pp. 542-551, 2014.

[5] Y. Wang, "A neural network adaptive control based on rapid learning method and its application," Advances In Modeling and Analysis, Vol. 46(3), pp. 27-34,199

[6] J Tang,Xia Hu,Huiji Gao,Huan Liu,” Unsupervised Feature Selection for Multi-

View Data in Social Media",Proceedings of the ACM international conference on Information and Knowledge management (2015),pp.1673-1678.

[7] Haytham Elzhazel and Alex Auseem, "Unsupervised Feature Selection with ensemble learning ", Machine Learning ,Springer (2013).

[8] Mohak Shah, Mario Marchand, and Jacques Corbeil, " Feature Selection with Conjunctions of Decision Stumps and Learning ", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 34, NO. 1,(2012) pp.174-186.

[9] Qingzheng XU, Lei WANG, Baomin HE, Na WANG," Modified Opposition-Based Differential Evolution for Function Optimization", Journal of Computational Information Systems 7:5 (2011) pp.1582-1591.

[10] Hyndman, R. et al., 2008. Forecasting with Exponential Smoothing: The State Space Approach, Springer Science & Business Media.