**WEATHER BASED CROP PREDICTION USING MACHINE LEARNING**

A Project Report submitted in partial fulfillment of the requirements for the award of the degree of

**Bachelor of Technology**

### in

**Computer Science and Engineering**

by

### Aman Mukati (112015011)

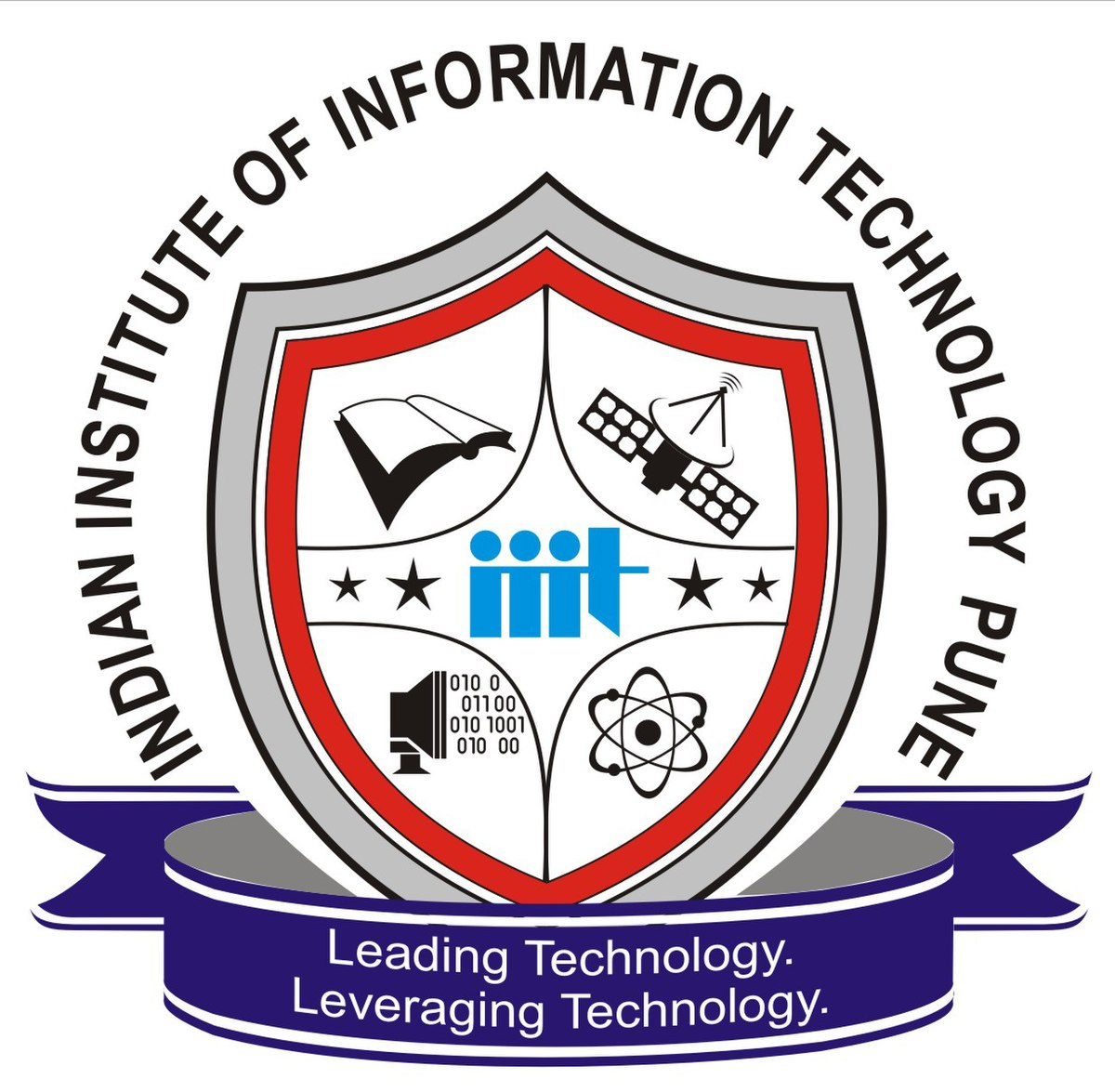
### Anishek Kumar Chaudhary (112015015)

### Harshal Patil (112015061)

### Pawan Bhawsar (112015101)

**Under the Supervision of: Tanmoy Hazra**

### Semester: V

****

#### Name of Department: Department of Computer Science and Engineering

#### Indian Institute of Information And Technology, Pune

**(An Institute of National Importance by an Act of Parliament)**

#### 

#### DECEMBER 2022

**BONAFIDE CERTIFICATE**

This is to certify that the project report entitled **“Weather Based Crop Prediction Using Machine Learning”** submitted by **Aman Mukati** bearing the **MIS No: 112015011**, **Anishek Kumar Chaudhary** bearing the **MIS No: 1120150015**, **Harshal Patil** bearing the **MIS No: 112015061**, **Pawan Bhawsar** bearing the **MIS No: 112015101** in completion of his/her project work under the guidance of **Tonmay Hazra** is accepted for the project report submission in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in the **Department of Computer Science and Engineering**, Indian Institute of Information Technology, Pune (IIIT Pune), during the academic year **2022-23**.

#### Tonmoy Hazra Tonmoy Hazra

Project Guide Head of the Department

Assistant Professor Assistant Professor

Department of CSE Department of CSE

IIIT Pune IIIT Pune

Project Viva-voce held on 15th December 2022

**Undertaking for Plagiarism**

We **Aman Mukati** bearing the **MIS No: 112015011**, **Anishek Kumar Chaudhary** bearing the **MIS No: 1120150015**, **Harshal Patil** bearing the **MIS No: 112015061**, **Pawan Bhawsar** bearing the **MIS No: 112015101** solemnly declare that research work presented in the **report/dissertation** titled **“Weather Based Crop Prediction Using Machine Learning”** is solely **our** research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete report/dissertation has been written by **us**. I understand the zero tolerance policy of **Indian Institute of Information Technology Pune** towards plagiarism. Therefore **We** declare that no portion of my **report/dissertation** has been plagiarized and any material used as reference is properly referred/cited. I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of the degree, the Institute reserves the rights to withdraw/revoke my **B.Tech** degree.

**Students’/ Student’s Name and Signature with Date**

**Aman Mukati : \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Anishek Kumar Chaudhary : \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Harshal Patil : \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Pawan Bhawsar : \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Conflict of Interest**

**Manuscript title:**

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers’ bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

**Students’/ Student’s Name and Signature with Date**

**Aman Mukati : \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Anishek Kumar Chaudhary : \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Harshal Patil : \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Pawan Bhawsar : \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Weather Based Crop Prediction Using Machine Learning

1. **Problem Statement.**

**“WEATHER BASED CROP PREDICTION USING MACHINE LEARNING”**

In this problem we are trying to solve the problems faced by the farmers due to adverse weather condition.

We are trying to optimize the crop production by informing farmers right time to plant the crops and which crop to plant for the given future expected weather condition using machine learning to predict it.

Problems such as extreme weather conditions, like dry spell, heavy rainfall, floods, and drought are major problem for the farmers and destroy crops.

1. **Objective.**

The Objective is to solve problems faced by the farmers due to unexpected weather. For that we are employing two methods

1. Selecting right crop for the given weather condition: We are using machine learning to select the right crop to be planted at the right time to maximize the yield obtained from a given farm land.

If a farmer is going to plant rice, he will need right amount of rainfall to create good paddy to plant the rice in. If there is no rain for next 3 days, the farmer can wait for 3 days before getting saplings and renting the equipment needed for planting rice. This will save his crops from damage due to lack of rain and also save on leasing cost of equipment.

1. Our other objective is to predict adverse condition which may affect the crops and warn the farmers so that they can take precautionary measures to reduce the effect of these adverse conditions.

## ACKNOWLEDGEMENT

This project would not have been possible without the help and cooperation of many. I would like to thank the people who helped me directly and indirectly in the completion of this project work.

First and foremost, I would like to express my gratitude to our honorable Director, **Prof. O.G. Kakde**, for providing his kind support in various aspects. I would like to express my gratitude to my project guide **Tanmoy Hazra**, **Department of CSE**, for providing excellent guidance, encouragement, inspiration, constant and timely support throughout this **M.Tech/B.Tech Project**. I would like to express my gratitude to the **Head of Department Tanmoy Hazra**, **Department of CSE**, for providing his kind support in various aspects. I would also like to thank all the faculty members in the **Department of CSE/ECE** and my classmates for their steadfast and strong support and engagement with this project.

Title

## Abstract

Farmers in India face many problems due to lack of information available about weather conditions which can be expected and best time to plant the crop.

India is predominantly an agricultural country with 55% of population in employed in agricultural sector. Though only 16% of India's 2.62 lakh crores GDP is contributed by the agriculture. There is a lot of hope for improvement.

To help mediate this problem we are developing a software to predict the best crops which the farmer can plant to optimize the yield which he receives form a given piece of land.

To mitigate the problems, first we need to identify the problems which are faced by the farmers.

Some of the problems which are faced by the farmers are:

1) Small and fragmented land holding.

2) Quality of seed

3) Reducing quality of soil due to lack of replenishment of nutrients.

4) Destruction of crops due to sudden weather change.

5) Reduced Yield due to delayed or early sowing.

Plating the right crop based on the conditions of the soil and weather can help with reducing the effect of the problems faced by the farmers.

Our system first predicts the weather based on the given past data of the weather and then using that data we have trained a weather prediction model which will predict the weather.

After we have predicted the weather, we take that data and use it to give list of best crops which can be planted on a particular land based on factors like predicted weather, current soil conditions which includes values like soil Ph level, its nitrogen, potassium, and phosphorus.t

**Keywords:** Weather Prediction, Machine Learning, Crop Prediction.

## TABLE OF CONTENTS

**Abstract i**

**(i)** [**List of Figures/Symbols/Nomenclature iv**](#_heading=h.gjdgxs)

**(ii)** [**List of Tables v**](#_heading=h.30j0zll)

1. **Introduction 1**

[1.1 Overview of work . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1](#_heading=h.1fob9te)

[1.2 Motivation of work . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1](#_heading=h.3znysh7)

[1.3 Literature Review. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1](#_heading=h.3znysh7)

1.4 Research Gap. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1

1. **Problem Statement**

2[.1 Research Objectives . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1](#_heading=h.3znysh7)

2[.2 Methodology of work . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1](#_heading=h.3znysh7)

1. **Analysis And Design 5**
2. **Results and Discussion 9**
3. **Conclusion and Future Scope 10**
4. **References**
5. **Publication (if any)**

# List of Figures [/ Symbols/ Nomenclature](#_heading=h.gjdgxs)

# List of Tables

**Chapter 1**

**Introduction**

## Overview of Work

Every industry is focusing on optimizing the work and productivity, it only makes sense that we also try to make the agriculture industry as optimized as possible.

To do this we are using machine learning to make a model which first predicts the weather and then it used that predicted data and other factors data which affect the crops to predict which crop is the best one to plant at the given location at that particular time.

Using our predictive model farmers will have a list of crops which he can plant at that particular time in those particular locations. Now farmer can select the crop which he thinks is viable financially and technically.

## Motivation of the Work

Factors like increase in population, decrease in total available area for agriculture, climate change, decrease in soil fertility, fest infestation, etc. It is or top most importance that we increase the yield of

* 1. [**Literature Review**](#_heading=h.3znysh7)

1. Machine Learning convergence for weather based crop selection

In This paper we studied about how machine learning can be used to select crop based on time and area so that we can maximize the crop yield. Owing to increasing population, the demand for crop production is increasing. On the other hand, the available land area for agriculture is decreasing day by day. It also suggests the proper sowing time for suitable crops using seasonal weather forecasting. Machine learning algorithms such as Artificial neural network is used for weather prediction, and Random forest classification algorithm is used to select suitable crops.

1. Big data analytics in Agriculture

This paper gives idea about how big data can be used in agriculture to get the quality and related data to be used in various business. The sole focus is to extract acumens from it which can be utilized by the farmers or the end users and can be implemented to gain and achieve assured outcomes. Such as Apt crop forecasting, precision farming, smart agriculture, achieving high quality seeds, climate predictions and much more. Big data has an enormous potential in the field of agriculture. The detailed understanding of Big data can help us overcome the significant challenges for any kind of traditional farming techniques that can lead to an increase in crop quality and production. Therefore, this technology in agriculture can be used to collect and analyze huge data that can be obtained from numerous stages in agriculture The major influence of implementing big data in agriculture are benchmarking, analytics, model prediction, visualization, marketing and management.

1. Applications of Machine Learning Techniques in Agricultural Crop Production

This paper has been prepared as an effort to reassess the research studies on the relevance of machine learning techniques in the domain of agricultural crop production. Methods/Statistical Analysis: This method is a new approach for production of agricultural crop management. Accurate and timely forecasts of crop production are necessary for important policy decisions like import-export, pricing marketing distribution etc. which are issued by the directorate of economics and statistics. However one has understand that these prior estimates are not the objective estimates as these estimate requires lots of descriptive assessment based on many different qualitative factors. Hence there is a requirement to develop statistically sound objective prediction of crop production. That development in computing and information storage has provided large amount of data.

1. PREDICTION OF MONTHLY RAINFALL IN CHENNAI USING BACK PROPAGATION NEURAL NETWORK MODEL

This paper presents a new approach using an Artificial Neural Network technique to improve rainfall forecast performance. A real world case study was observed in Chennai for 32 years of monthly mean data with meteorological parameters such as wind speed, mean temperature, relative humidity, aerosol values (RSPM) in the area were used to develop the ANN model .In order to forecast rainfall in Chennai City, Back Propagation Neural Networks (BPNNs), a data driven technique based on the working principle of biological neurons are applied in this study. The mean monthly rainfall is predicted by using ANN model. The model can perform well both in training and independent periods.

1. WB-CPI: Weather Based Crop Prediction in India Using Big Data Analytics

This paper aims at collecting and analysing temperature, rainfall, soil, seed, crop production, humidity and wind speed data (in a few regions), which will help the farmers improve the produce of their crops. Firstly, we pre-process the data in a Python environment and then apply the MapReduce framework, which further analyses and processes the large volume of data. Secondly, k-means clustering is employed on results gained from MapReduce and provides a mean result on the data in terms of accuracy. After that, we use bar graphs and scatter plots to study the relationship between the crop, rainfall, temperature, soil and seed type of two regions (Ahmednagar, Maharashtra and, Andaman and Nicobar Islands). Further, a self-designed recommender system has been used to predict the crops and display them on a Graphic User Interface designed in a Flask environment. The system design is scalable and can be used to find the recommended crops of other states in a similar manner in the future.

**1.4 Research Gap**

After reading above research paper we get the idea about how machine learning and it’s algorithms and BIG DATA can be used in agriculture so that we can get more crop production. And we found that most of these paper get the good and precise weather data. But to train the model they chose they used limited data for limited 1-2 cities and most of them forgot to consider the important soil factors that are required in order to select a perfect crop and to get good crop production.

So we considered the data of 3 different cities of India where the variability of rain is so much more than other cities of India like Pune, Bangalore, Hyderabad. And we collected the 11 years of detailed weather data from 2009-2020 so our model can learn from that and give us good results. And another factor we considered is soil factors like Nitrogen content , Phosphorus content and Potassium content in the soil, soil PH, soil temperature, soil humidity and based on that we found the suitable crops.

# Chapter 2

# Problem Statement

The biggest motivation of doing this paper is Now a days X-ray and CT in medicine use high-resolution images, they require high specification equipment and huge energy consumption due to high computation in learning and recognition, incurring huge costs to create an environment for operation. Thus, this paper proposes a chest X-ray outlier detection model which will decrease the working time.

**2.1. Research Objectives**

**2.2. Methodology of the Work**

# Chapter 3

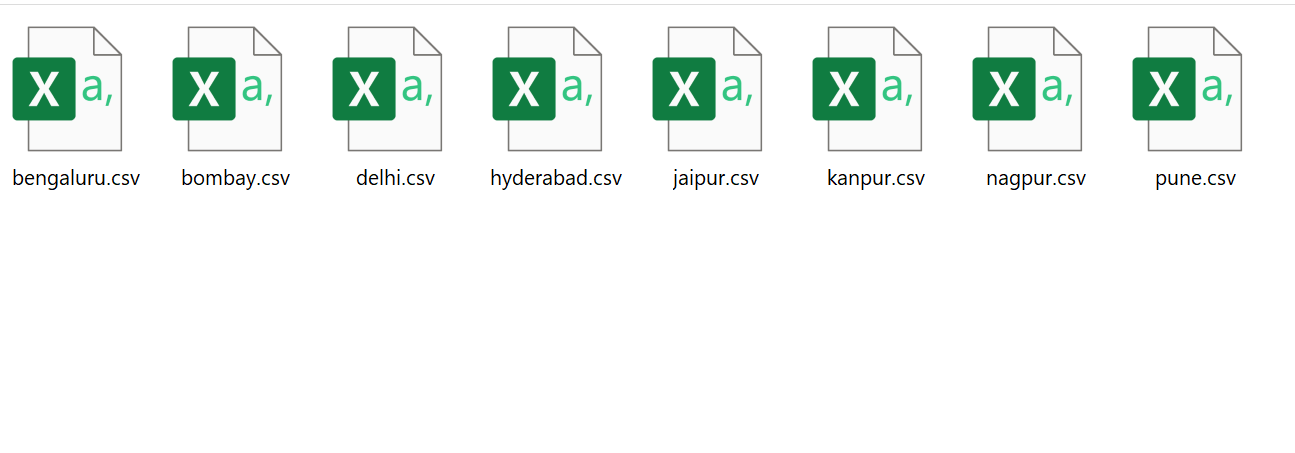
# Analysis and Design

In This part we will focus on how to design the model where to start doing work and after that we will analyze it.

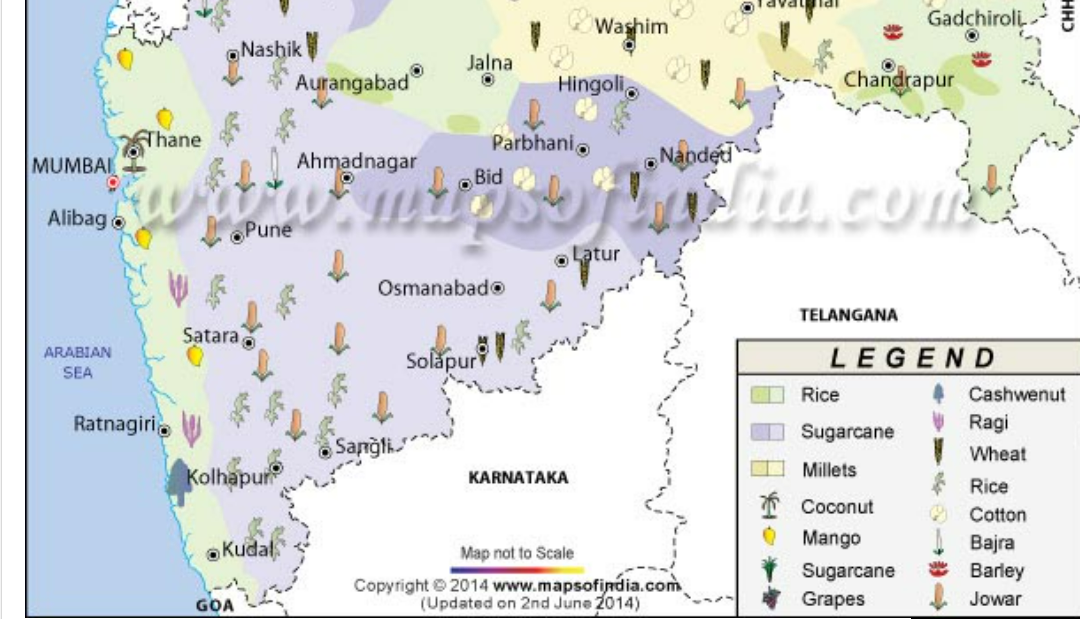
So first part will be data collection and data preprocessing so that information that can be derived from it will be useful and our model will easily learn from that data.

1. **DATA COLLECTION**

Various data sets were collected during this step. Facing a little difficulty we found the hourly data of 8 major cities from 2009-2020 : BANGALORE, MUMBAI, PUNE, JAIPUR, DELHI, HYDERABAD, NAGPUR, KANPUR (Figure-1) from Kaggle and government websites(Figure-2). Then we selected 3 main cities where rainfall and climate change is so variable like PUNE, BANGALORE, HYDERABAD. After getting the weather data we gone through some state government websites and found the crops which have high production in that area. And after searching days we get the soil data from government website in the unstructured form in text and small tables which consists of soil moisture, soil temperature in that area.



Figure



Figure

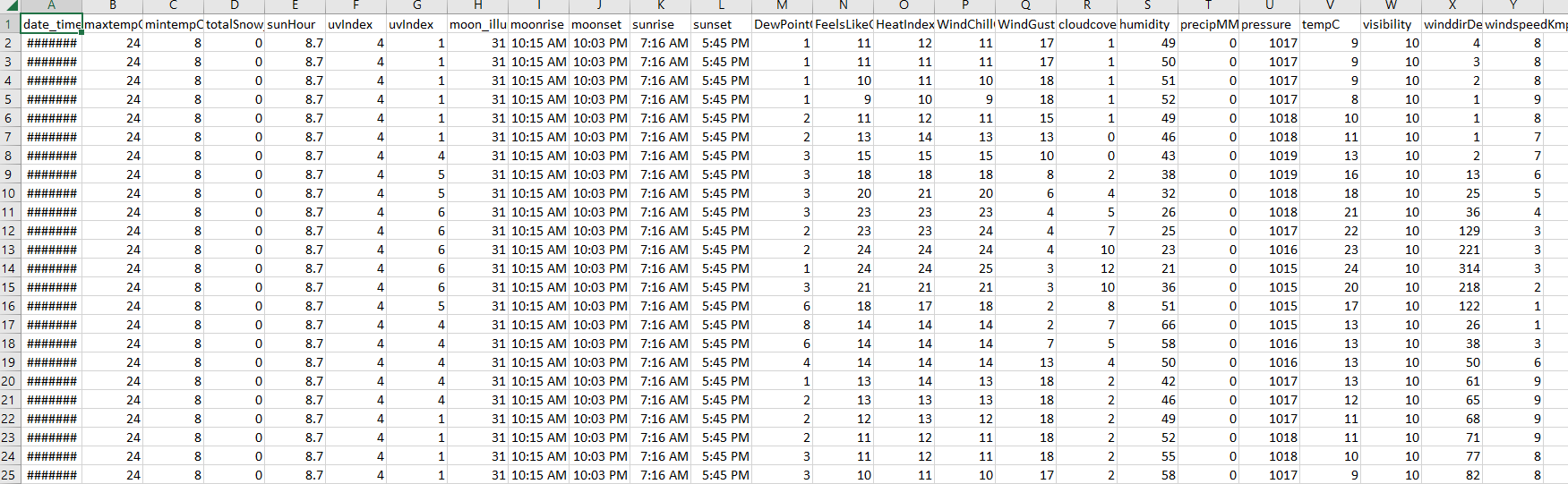
1. **DATA PREPROCESSING**

After collecting data from various government websites we need to prepare or preprocess the data so that we will use that in our model. So as we have weather data of 8 cities we will choose 3 cities and will combine their data. And after that we will add a column which will show us the city 0 for pune,1 for bangalore, 2 for Hyderabad, then we will remove correlated data which is composite and derived and we made columns which contains information in string like time : so we separated that column into different columns like day | month | year | hour| .

Then we have soil data so we searched soil conditions and climate conditions of these cities and chose which crops will be suitable for that area. And we made a csv file which has a crop and related conditions according to that crop like how much temperature and rainfall will be required and after making that table we moved again to final weather dataset and we will find the sum of rainfall, monthly rainfall to find total rainfall in that month and that year and the average of humidity and temperature.

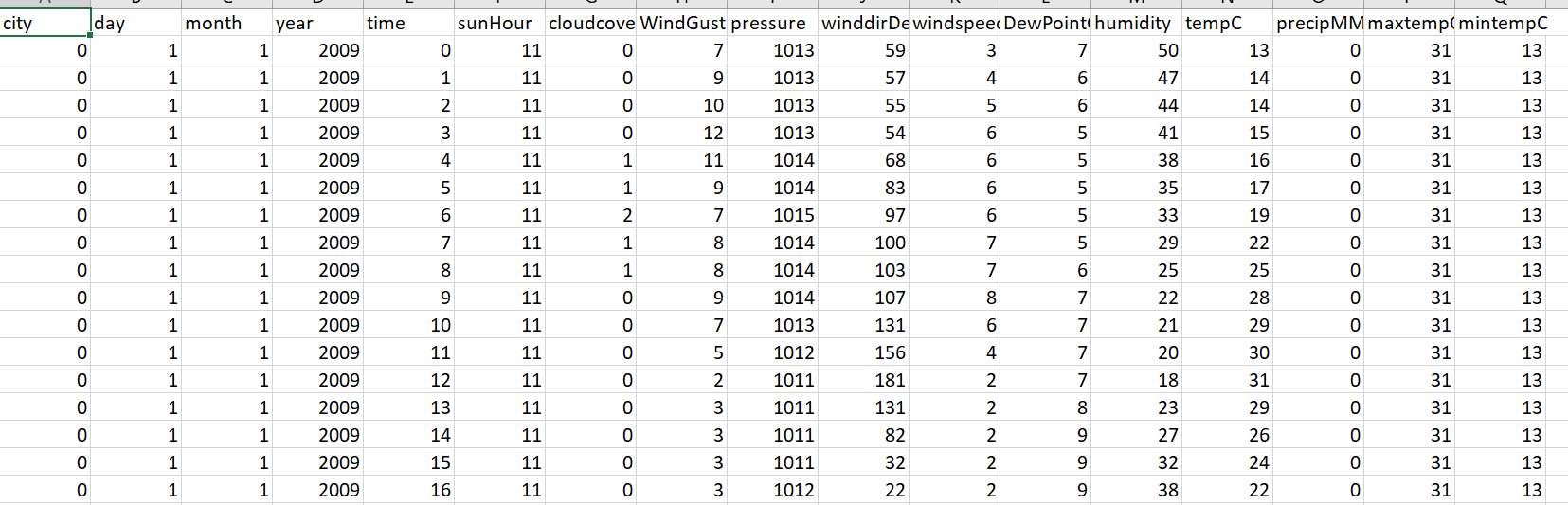
Here we will extract some features like humidity, temperature, rainfall in that city and combined them and form the different csv file so we can compare and find crops. Which are suitable for that.

Before preprocessing data looks like :- unorganized weather data.



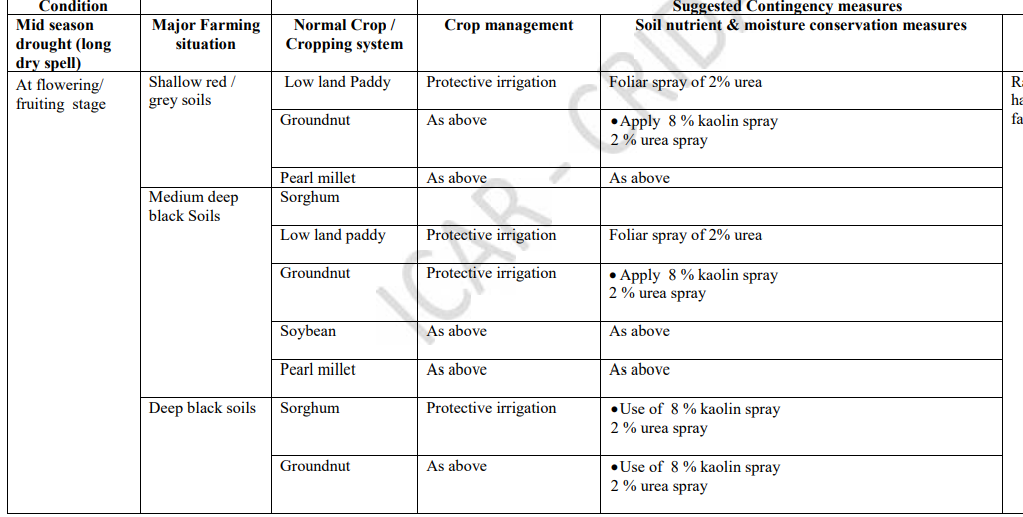
Figure

After preprocessing weather data looks like :- organized and well relatable.

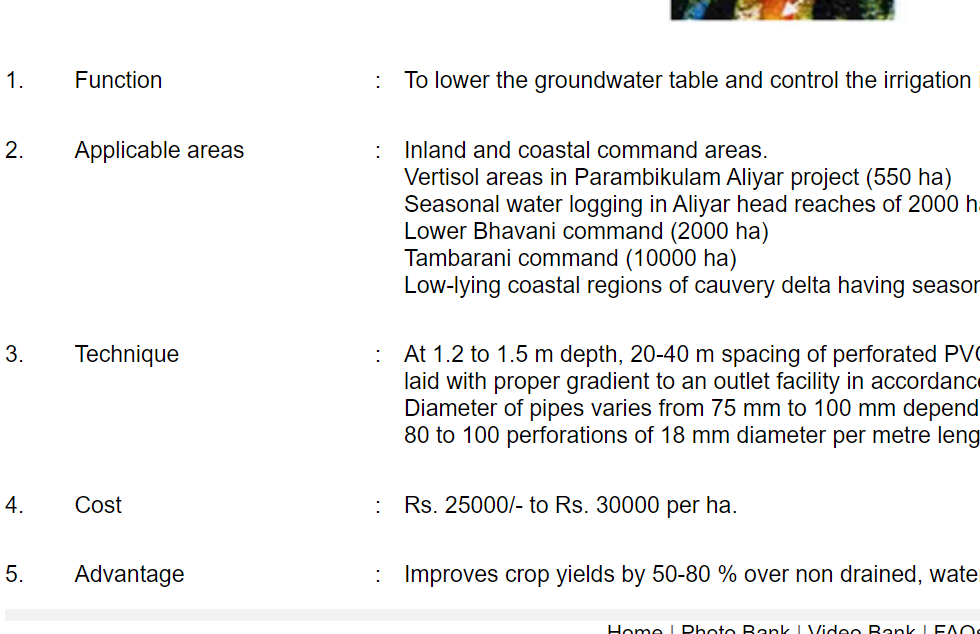


Figure

Then we have soil data which is completely written in sentences and unstructured.

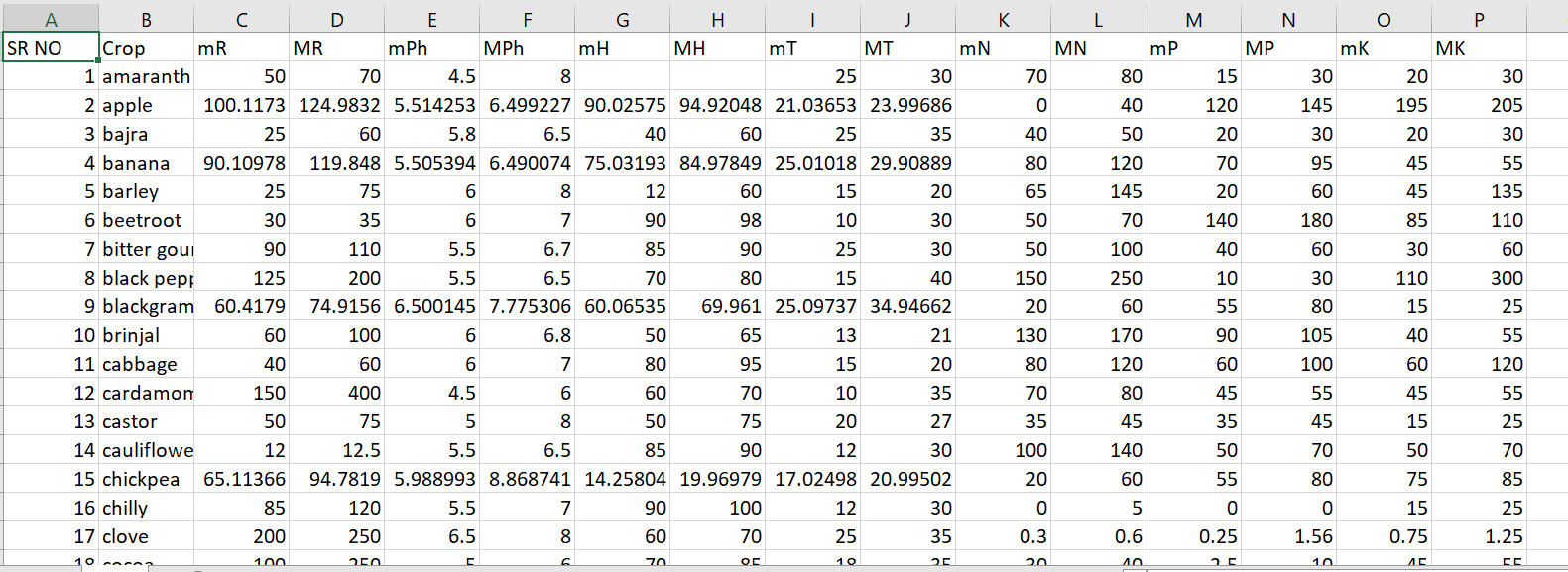


Figure

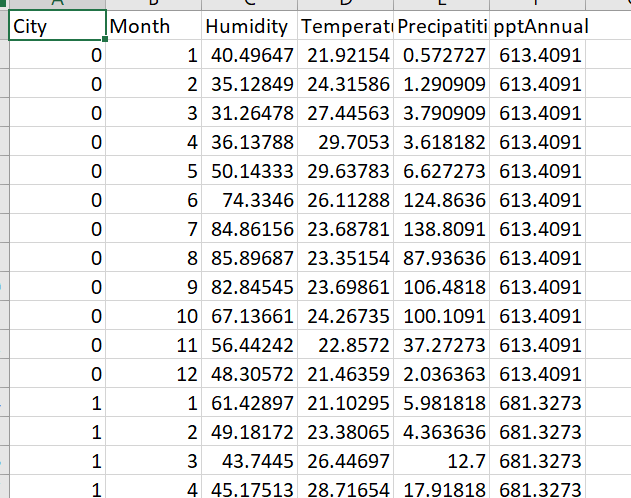


Figure

After preprocessing we have a structured data in csv file.



Figure



Figure

1. **MODEL TRAINING**

Now after collecting and preprocessing that we need to choose the machine learning algorithm which will best fit the data.

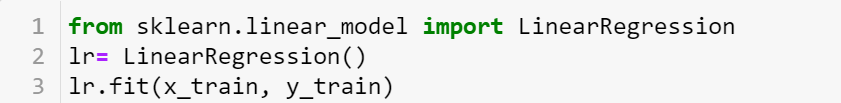
As first we need to predict the weather then based on that weather, we will choose the crop which is suitable.

* 1. **WEATHER PREDICTION**

For the weather prediction we need to predict the humidity and rainfall and temperature.

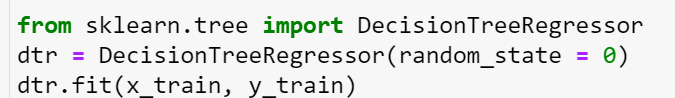
So it’s a regression type problem where we need to find the output which is continuous and we have 4 best algorithms

1. Multiple Linear Regression



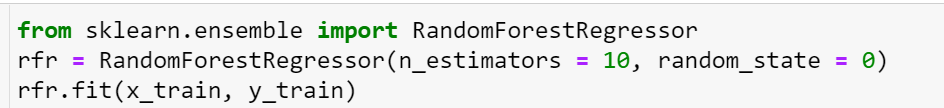
Figure

1. Decision Tree Regression



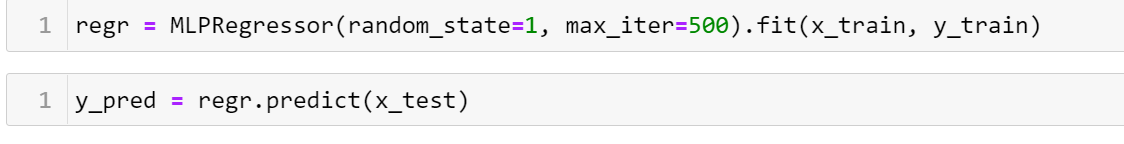
Figure

1. Random Forest Regression



Figure

1. Artificial Neural Network



Figure

Crop selection model takes input of soil parameters such as soil temperature, soil pH, soil humidity or moisture, soil fertility (N,P,K) and its water holding capacity. The soil and predicted weather parameters are used collectively to choose suitable crops for land. The crop selection can be done either annual or seasonal. In seasonal crop selection method, the model suggests one or more crop that is suitable for a season and also suggests crop requirement such as irrigation, and the proper sowing time of a particular crop based on predicted weather.

The dataset consists of crops which are rice, cotton, maize, sunflower, castor, chilli, redgram, greegram, jowar and soyabean etc... However, the model can be extended for different type of land and crops. A random forest classification method is used to classify crops based on weather and soil parameters. A random forest classifier is an ensemble classifier which uses collection of decision trees created from subset of training dataset. The outcome of random forest is decided by aggregating output of each decision tree. The class having maximum votes after aggregation becomes the outcome of random forest.

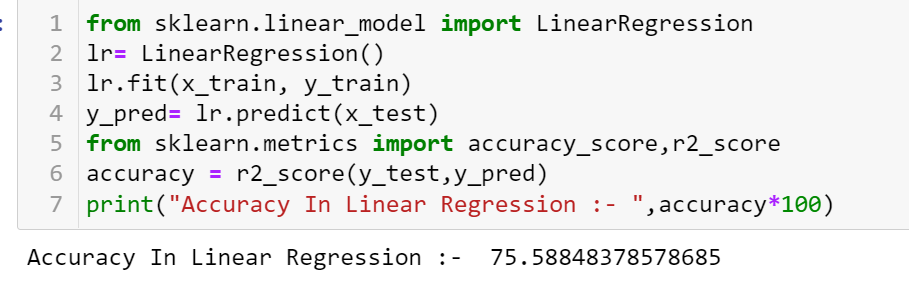
After that we will find suitable crops based on the location and the month.

Figure

# Chapter 4

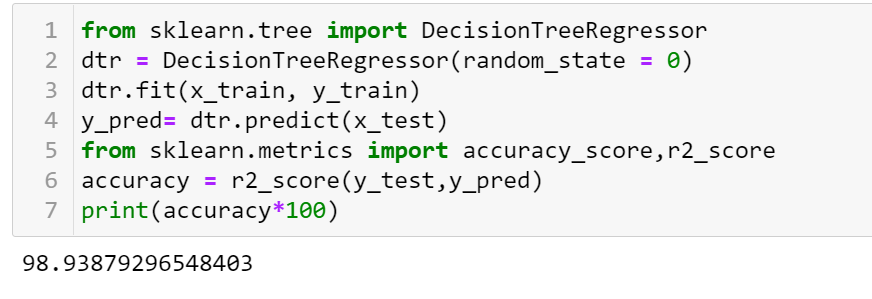
# Results and Discussion

So we in weather data we first applied linear regression and in that we got the accuracy of 75%.



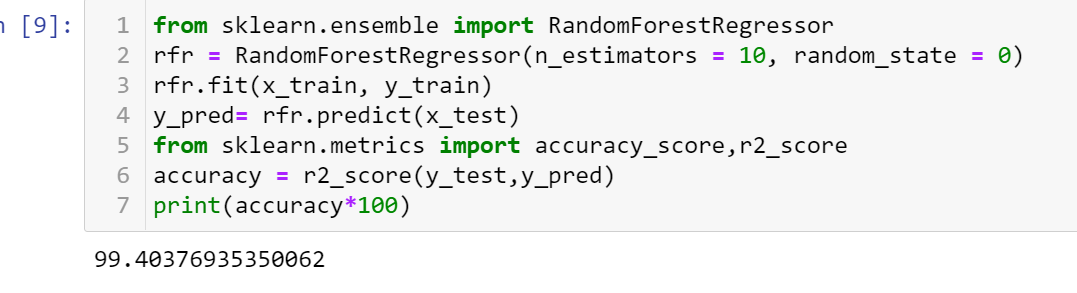
Figure

Then we will apply decision tree to get more accuracy and using decision tree we got the accuracy of 98% :-



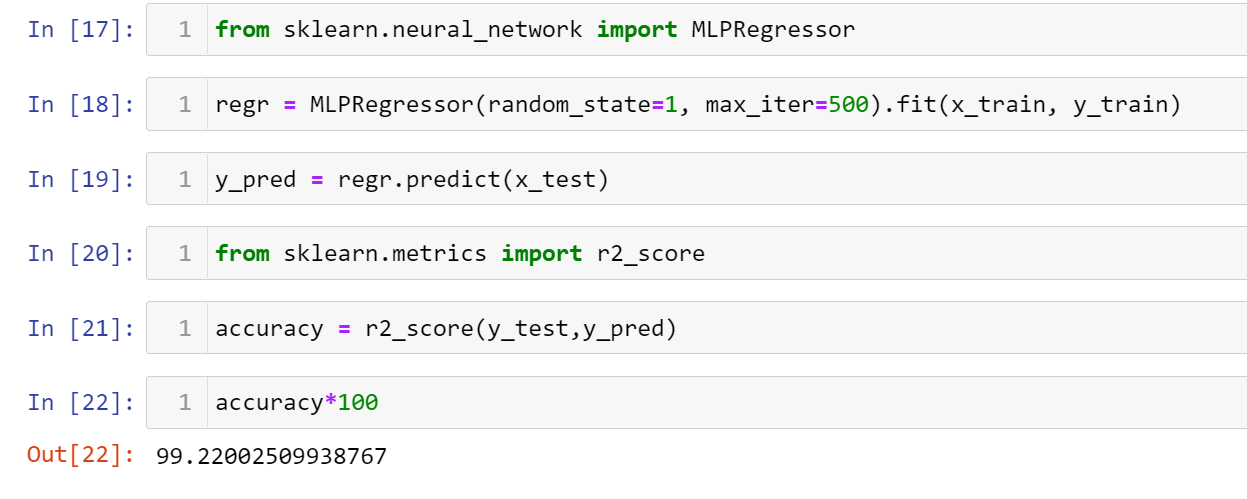
Figure

After applying decision tree we thinked about random forest as it can work with large data and will combine results of various decision trees so using random forest we got the accuracy of 99% :-



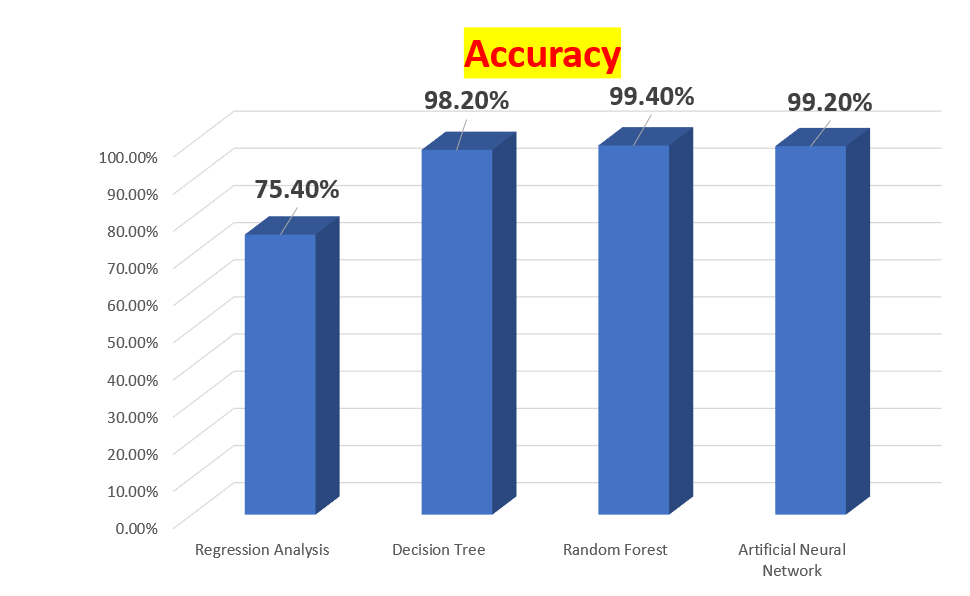
Figure

After Random Forest we applied Artificial Neural Network and then we got the accuracy of 99.2% :-



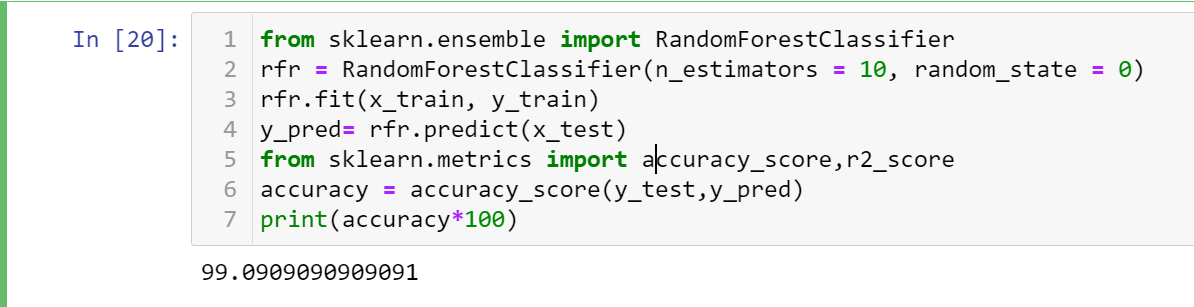
Figure

So we can see that the random forest and artificial neural network works best in this case so we can choose ANN for the prediction of weather.



Figure

And for the crop selection we will use random forest which gave the accuracy of 99% in crop selection.

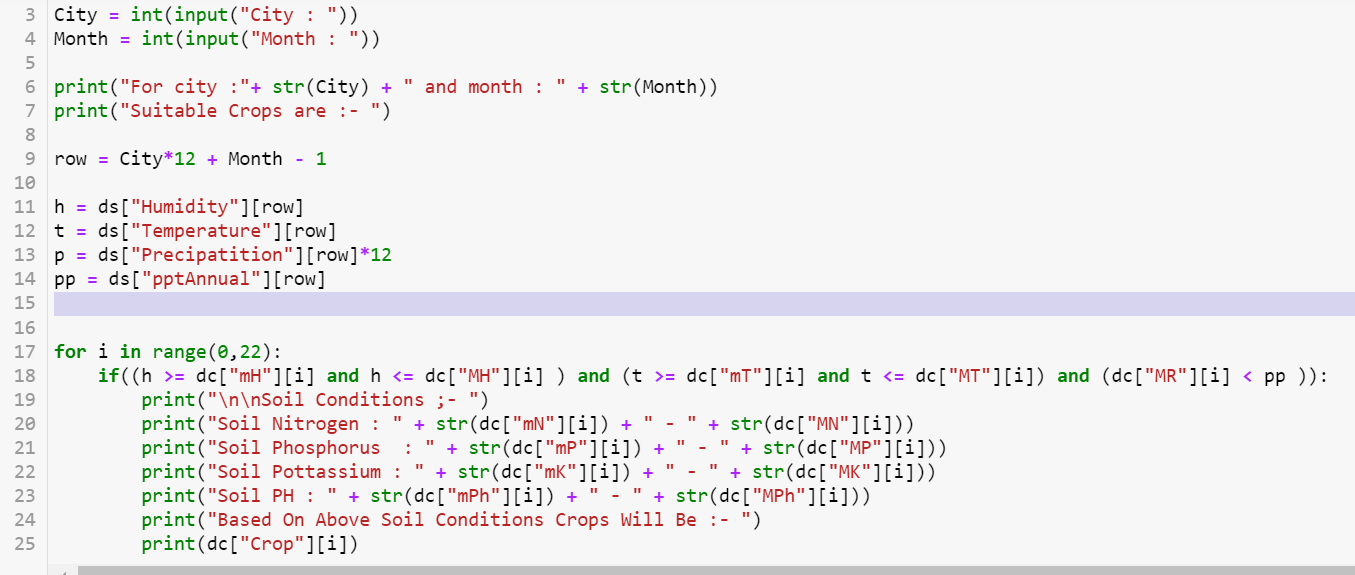


Figure

And ex. For city = pune - 0

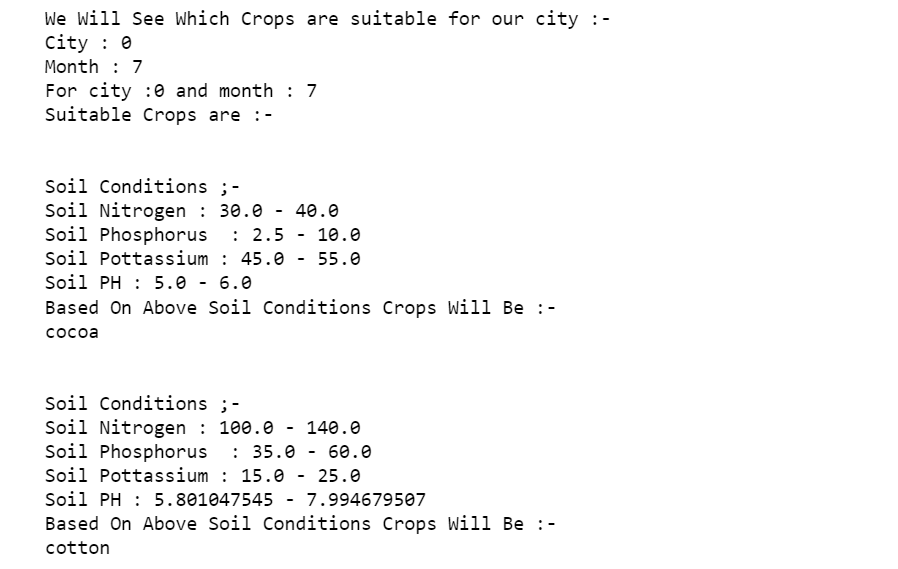
And month = July – 7

Which crops will be suitable;-



Figure

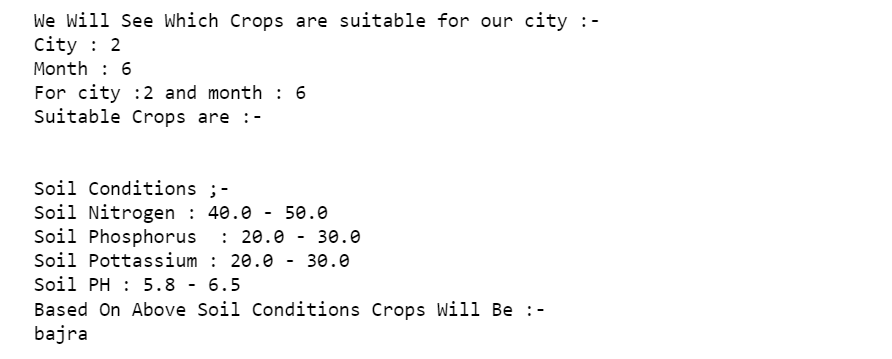
Results ==



Figure

City = Hyderabad

Month = June



Figure

# Chapter 5

# Conclusion and Future Scope

The proposed method represents a novel weather based crop selection system to select crops for land, based on predicted weather parameters and its soil parameters.

The classification model is trained by considering the weather and soil parameters of agro-climatic zones of Maharashtra Telangana and Karnataka . The seasonal weather forecasting is carried out using ANN. A random forest classifier is used to get more than one suitable crop for a land. Moreover, the proposed method also suggests the proper sowing time for each crop based upon predicted weather parameters.

This work can be improved if we will use more precise weather data collected from Advance IOT sensors. Factors like soil type, irrigation, cloud cover etc. may be included in the system to refine its output. And we can consider those crops which have high production in that area and we can use that and crop yield factor combined to get more accurate crops. Also, the recommender can be modified to warn about the diseases that can occur in a crop in a particular season and suggest the types of fertilizers or nutrients needed in the soil for the crop to grow and give its best yield.

# References

**Instructions**

1. All the pages should be numbered at the bottom right side of the page.
2. The formatting of dissertation/report should be in APA format.
3. All the References should be mentioned in the APA format.
4. All the tables and figures should be labeled. The compiled list of the tables and figures should be presented in the appendix.
5. The report should be original in its work. Only 20% of plagiarism is permissible. Students must attach an Undertaking for plagiarism report at the designated place of the report.
6. The students must get the report verified, evaluated and signed by the supervisor before its final submission in the spiral bound form and incorporate any changes, suggestions or recommendations as directed by the supervisor.
7. The students should bring the final report (spiral bound) on the day of End-Semester Project Presentation and submit it to their respective panelists.
8. The format of BTP/MTP report shared should be strictly followed.
9. The Supervisor-Student Interaction Report should be submitted to the panelists on the day of Viva-Voce Examination.
10. The Problem Statement and Objectives submitted during the Mid-Term Presentation must be attached to the Final report after the Bonafide certificate given in the sample report.
11. The **order of report** should be in the following format:
12. Title Page
13. Bonafide Certificate, Conflict of Interest and Undertaking for Plagiarism
14. Problem statement and objectives submitted during the Mid-Term Presentation
15. Abstract with keywords
16. Contents
17. Chapters as suggested in report

11. All the data/tables/figures taken from other sources should be duly credited in the footnotes or references.

12. The font size should be 12 throughout the document. The title should be Font size 17 and sub-heading should be Font Size 14, followed by under subheadings as Font Size 12. The Font Style must be **Times New Roman Only** and spacing between each line must be 1 and between headings should be 2.

13. The spacing between each line must be 1.15 and between headings must be 2.

14. All the paragraphs must be aligned as “justify” while the headings must be in center, the sub-headings should be left aligned.