**A Project Report**

**On**

# SOIL PREDICTION FOR SUITABLE CROPS USING KNN ALGORITM

*Submitted to*

**JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY ANANTAPUR, ANANTHAPURAMU**

In Partial Fulfillment of the Requirements for the Award of the Degree of

**BACHELOR OF TECHNOLOGY**

**In**

**COMPUTER SCIENCE & TECHNOLOGY**

**Submitted By**


**P Vishnu Vardhan Naidu - (19691A28H9)**

**CR Tejavardhan Reddy - (19691A28F8)**


**Under the Guidance of**

**Mr. D Abdul Jaleel**

**Assistant Professor,**

**Department of Computer Science & Technology**



**MADANAPALLE INSTITUTE OF TECHNOLGY & SCIENCE**

**(UGC – AUTONOMOUS)**

**(Affiliated to JNTUA, Ananthapuramu)**

**Accredited by NBA, Approved by AICTE, New Delhi)**

**AN ISO 9001:2008 Certified Institution**

**P. B. No: 14, Angallu, Madanapalle, Annamayya–517325, 2019-2023**

✆ : 08571 - 280255, 280590
MADANAPALLE INSTITUTE OF TECHNOLOGY & SCIENCE
An Engineering College Sponsored by :
RATAKONDA RANGA REDDY EDUCATIONAL ACADEMY
Approved by AICTE, New Delhi and Affiliated to JNTUA, Anantapur
P.B. No. 14, Angallu, MADANAPALLE - 517 325. Chittoor Dist. A.P.
www.mits.edu

**DEPARTMENT OF COMPUTER SCIENCE & TECHNOLOGY**

## BONAFIDE CERTIFICATE

This is to certify that the project work entitled "**SOIL PREDICTION FOR SUITABLE CROP USING KNN ALGORITHM**" is a bonafide work carried out by

**P Vishnu VardhanNaidu - (19691A28H9)**

**CR Tejavardhan Reddy - (19691A28F8)**

Submitted in partial fulfillment of the requirements for the award of degree **Bachelor of Technology** in the stream of **Computer Science & Technology** in **Madanapalle Institute of Technology and Science, Madanapalle,** affiliated to **Jawaharlal Nehru Technological University Anantapur, Ananthapuramu** during the academic year 2022-2023

SUPERVISOR

**Mr.D Abdul Jaleel**

**Associate Professor,**

**Department of CST**

Head of the Department

**Dr.M.Sreedevi**

**Professor and Head**

**Department of CST**

Submitted for the University examination held on:

**Internal Examiner**

**Date:**

**External Examiner**

**Date**

## ACKNOWLEDGEMENT

We sincerely thank the **MANAGEMENT** of **Madanapalle Institute of Technology and Science** for providing excellent infrastructure and lab facilities that helped me to complete this project.

We sincerely thank **Dr. C. Yuvaraj, M.E., Ph.D**., **Principal** for guiding and providing facilities for the successful completion of our project at **Madanapalle Institute of Technology and Science,** Madanapalle.

We express our deep sense of gratitude to **Dr. M. Sreedevi, MTech., Ph.D., Professor and Head of the Department of CST** for her continuous support in making necessary arrangements for the successful completion of the Project.

.

We express my deep sense gratitude to **Dr.Manikandan R,M.E., Ph.D., Project Coordinator,** for his guidance and encouragement that helped us to complete this project.

We express our deep gratitude to my guide **Mr.D Abdul Jaleel, Assistant Professor, Department of CST** for his guidance and encouragement that helped us to complete this project.

We also wish to place on record my gratefulness to other **Faculty of CST Department** and to our friends and our parents for their help and cooperation during our project work.

MADANAPALLE INSTITUTE OF TECHNOLOGY & SCIENCE
(UGC-AUTONOMOUS)
Estd 1998
Affiliated to JNTUA, Anantapur & Approved by AICTE, New Delhi
Recognised Research Center
Accredited by NBA for CSE, ECE, EEE & ME
World Bank funded Institute
Recognised by UGC under the sections 2(f) and 12(B) of the UGC act 1956
Recognised as Scientific & Industrial Research Organization by DSIR of DST

**RECOGNISED RESEARCH CENTER**

***Plagiarism Verification Certificate***

This is to certify that the B.Tech Project report titled, "**Soil Prediction for suitable crop Using KNN Algorithm**" submitted by **P.VishnuVardhan Naidu (19691A28H9),CR.Tejavardhan Reddy (19691A28F8)** has been evaluated using **Anti- Plagiarism Software, TURNITIN** and based on the analysis report generated by the software, the report's similarity index is found to be 10 % .

**The following is the turnitin report for the project report consisting of 60 pages.**

# DECLARATION

We hereby declare that the results embodied in this project **"Soil Prediction for suitable crop Using KNN Algorithm"** by us under the guidance of **Mr.D.Abdul Jaleel Assistant Professor, Department of CST** in partial fulfillment of the award of **Bachelors of Technology** in **Computer Science & Technology** from **Jawaharlal Nehru Technological University Anantapur, Anantapur.** and we have not submitted the same to any other University/institute for award of any other degree.

**Date:**

**Place:**

**PROJECT ASSOCIATES**

**P Vishnu Vardhan Naidu**

**CR Tejavardhan Reddy**

I certify that the above statement made by the students is correct to the best of my knowledge.

**Date:** **Guide :**

## TABLE OF CONTENT

## List of Figures

## LIST OF ABBREVATIONS

| | |
|---|---|
| **KNN** | K-nearest Algorithm |
| **ML** | Machine Learning |
| **DT** | Decision Tree |

# ABSTRACT

The project is used to discover the best crop prediction model, which helps the farmers to decide the suitable crop to cultivate based on the available percentage of nutrients in the soil. As India is one of the top Agricultural producing Countries in the World and we are highly relay on agriculture. There are three important nutrients in every soil, which are called as the macro nutrients: Nitrogen (N), Phosphorus (P), and Potassium(K). In this project, we will use K-Nearest Neighbour algorithm. Here the N:P:K values plays a vital role to give the prediction of suitable crops for that particular soil. Machine learning is a decision support tool for crop prediction, including supporting making decisions on what crop to grow in the soil.

# CHAPTER 1
# INTRODUCTION

For most non-industrial nations, agribusiness is their essential wellspring of income. Present day horticulture is a continually developing methodology for rural advances and cultivating procedures. It becomes provoking for the ranchers to fulfill our planet's developing prerequisites and the assumptions for vendors, clients, and so on. A portion of the difficulties the ranchers face is-

i. Supplement lack in the dirt, brought about by a deficiency of vital minerals like potassium, nitrogen, and phosphorus can bring about diminished crop development.
ii. (ii) Ranchers commit an error by developing similar harvests for many years without trying different things with various assortments. Agribusiness is the main stockpile of Indian Economy. For the better harvest yield, the ranchers generally require a right yield that can give great yield in that dirt. It is anticipating future potential harvests that can be planted in soil with its separate NPK rates.

## 1.1 MOTIVATION

As we are seeing numerous ranchers committing Suicides these days by developing harvests without knowing the appropriateness of the yield. By taking into account this situation, we dealt with an undertaking which proposed reasonable yield for the dirt. Ranchers are relocating to urban communities in the wake of dealing with numerous issues. Legitimate Use of the accessible assets are getting decreased and most concerning issues looked by ranchers is because of absence of accessibility of Composts, manures, biocides, seeds, apparatuses not with regards to raising harvests.

## 1.2 PROBLEM DEFINITION

Without knowing the N,P,K values of particular crop that grows in the soil, Farmer grow the crop which results in decrease of crop production and crop yield. So in this project, We will find the proper N,P,K values of the soil .Based on that values we suggest the crop for that particular soil ,So that the crop yield and crop production increases.

## 1.3 OBJECTIVE OF THE PROJECT

- The primary goal of this study is to examine, evaluate and identify the suitable crop using trained models present in the dataset.
- And this model increases the accuracy in detecting the prediction of crop by using KNN Algorithm.

## 1.4 LIMITATIONS OF PROJECT

- lack of practical knowledge of farmers
- Here the detection of NPK will be done. The parameters like Pressure,humidity, PH were not identified.

## 1.5 ORGANIZATION OF THE DOCUMENT

- Chapter 2 is Literature Survey which emphasizes the related works and their disadvantages.
- Chapter 3 deals with Requirements and Analysis.
- Chapter 4 deals with the design.
- Chapter 5 deals with the Implementation and result of the Project.
- Chapter 6 deals with Testing and Validation.
- Chapter 7 deals with the Conclusion and Upcoming work.

# CHAPTER 2
# LITERATURE REVIEW

## 2.1 INTRODUCTION

To decrease crop failures and predict the best crops for farmers by using the latest technologies to give the best accuracy rates and good crop yielding we need to go through the different scenarios and test cases to give the best outcome.

## 2.2 LITERATURE SURVEY

In the research conducted by the Nischitha k, DhanushVishwakarma, Mahindra N, Ashwini , Manju Raju M.R " **Crop Prediction System using Machine Learning Approaches** [1]" **.** As we probably are aware the way that, India is the 2nd biggest populace country on the planet and larger part of individuals in India have agribusiness as their occupation. Ranchers are developing same harvests more than once without attempting new verity of yields and they applying composts in irregular amount without the lacking substance and amount. Thus, this is straightforwardly influencing on crop yield and furthermore causes the dirt fermentation and harms the top layer. In this way, we have planned the framework utilizing AI calculations for advancement of ranchers. Our framework will propose the best appropriate harvest for specific land in view of content and climate boundaries. And furthermore, the framework gives data about the expected substance and amount of composts, required seeds for development. Consequently, by using our framework ranchers can develop another assortment of harvest, may increment in net revenue and can stay away from soil contamination.

In the research conducted by the Mayank Chempaneri, Chaitanya Chandvidhkar, Darpan Chachpara,MansingRathod "**Crop Yield Prediction using Machine Learning**[2]".The impact of natural change in India, most of the cultivating yields are overall genuinely influenced concerning their show throughout a period of the new 20 years. Predicting the reap yield before of its accumulate would serve to the system that makers additionally, farmers for taking reasonable measures for advancing and amassing. The endeavor will In India, there are various approaches to extending the monetary improvement in the field of agribusiness. Data digging is similarly important for expecting crop yield creation. By the huge, data mine is the technique associated with researching data according to alternate point of view and summarize it into

critical information. Unpredictable forest area is the most renowned and solid controlled estimation fit f0r perf0rming both portrayal and backslide tasks, that w0rk by build1ng countless decision trees during planning time and making result 0f the class that 1s the technique 0f the class (portrayal) mean assumption for the single trees.

In the research conducted by the S.P raja, Barbara Sawicka, Zoran Stankovic **"Crop Prediction Based on the Agricultural Environment[3]"**. Farming is a developing field of exploration. Specifically, crop expectation in agribusiness is basic also, is chiefly dependent upon soil and climatic conditions, includes precipitation, mugginess, and temp. Previously, ranchers had the option to settle on the harvest to be developed, screen its development, and decide when it very well may be gathered. Today, in any case, quick changes in natural circumstances have made it difficult for the cultivating local area to keep on doing as such. Subsequently, lately, AI procedures have assumed control over the errand of expectation, and this utilized a few of these to decide crop yield. To guarantee that AI (ML) model works at an elevated degree of accuracy, it is basic to utilize efficient highlight choice strategies to processor the crude information into an effectively processable AI agreeable data. To decrease more exact, just information includes that have a level of significance in deciding the result of the model should be utilized. Consequently, ideal component choice emerges to guarantee that main the most important highlights are acknowledged as a piece of the model. Conglomerating each and every element from crude information without checking for their job during the time spent causing the model will to superfluously confuse our model. Moreover, extra elements which contribute essentially nothing to the ML model will expand now is the ideal and intricacy influence the precision of the model result. The outcomes portray that a group strategy offers preferable expectation exactness over the current classification procedure.

In the research conducted by the Saravana kumar Venkatesan,Jonghyun Lim, and Yongyun Cho **"Crop Growth Prediction Model Based on Smart Farms[4]".** In the new past, the

horticultural business has quickly digitalized as savvy ranches through the wide use of information examination and artificial knowledge. Normally, high working costs in ranch are fundamentally client energy use. Therefore, exact assessment of horticultural energy use and natural elements is viewed as one of the significant undertakings for development controls. e development successions of yields in rural conditions like savvy ranches are connected with horticultural energy utilization and utilization. is concentrate on expects to create and approve a calculation that can decipher the yield development rate reaction to natural and sun powered energy factors in view of AI, and to assess the calculation's precision thought about to the base model. e proposed still up in the air through a near trial of three delegate AI methods, which are irregular timberland, SVM , and slope helping machine, taking into account the energy utilization for ecological control is exceptional development. harsh the execution with genuine information assembled from shrewd ranch Korea, the staggered foresee development with a precision of 0.88, taking into account information investigation of variables that utilization sun powered energy.

In the research conducted by the S.Iniyan, V.Akhil Varma, Ch.Teja Naidu **"Crop yield prediction using machine learning** [5]**".**. A few calculations have been applied t0 help crop yield expectation research. In this review, we played out a Precise Writing Audit (SLR) to separate and combine the calculations and elements that have been utilized 1n crop yield expectation studies. In light of our hunt standards, we recovered 5_6_7 applicable examinations from 6 electronic data sets, of which we have chosen 50 examinations for additional examination utilizing consideration and rejection measures. We examined these chose concentrates cautiously, dissected the strategies and elements utilized, and gave ideas to additional exploration. As per our examination, the most utilized highlights are temperature, precipitation, and soil, and the most applied calculation is Counterfeit Brain Organizations in these models. After this perception in light of the analysis of Intelligence based 5o papers, we played out an extras pursuit in electronic data sets to distinguish profound learning-based examinations, arrived at 30 profound learning-based papers,& separated the applied profound learning calculations. As this extra investigation, Convolution Neural brains Organizations (CNN) is the most broadly ut1lized

pro-found learn1ng calculati0n in these examinations, and the other generally utilized profound learning calculations are Long-Momentary Mem0ry (LSTM) and Profound Brain Organizati0ns (DNN).

In the research conducted by the Thomas van Kompenburg a, Ayalew Kassahun a, Cagatay Catal **"Crop Yield Prediction using Machine Learning Algorithms**[6] **".** Agribusiness is factor which, first and foremost, is significant for endurance. AI (ML) could be an essential viewpoint for procuring genuine world and usable answer for crop yield issue. Taking into account the current framework including manual counting, environment shrewd bug the executives and satellite symbolism, the outcome got aren't exactly precise. This paper centers fundamentally around foreseeing the yield of the harvest by applying different AI methods. The classifier models utilized here incorporate Calculated Relapse, Innocent Bayes and Arbitrary Backwoods, out of which the Irregular Woodland gives most extreme exactness. The forecast made by AI calculations will assist the ranchers with coming to a choice which harvest to develop to incite the most yield by considering factors like temperature, precipitation, region, and so on. This overcomes any barrier among innovation and farming area.

In the research conducted by the Mahendra N , Dhanush Vishwakarma , Nischitha K , Ashwini, Manjuraju M. R **"Crop Prediction System using Machine Learning Approaches[7]"** One of the important occupations practised in India is horticulture. It is the largest monetary area and plays a major role in the advancement of the nation as a whole. To meet the needs of 1.3 billion people, more than 60% of the country's land is used for horticulture. Consequently, adopting new farming innovations is crucial. This will motivate our nation's ranchers to their advantage [1]. Based on ranchers' previous experience in a particular area, an earlier harvest estimate and yield assumption was made. For their region, they will tend to favour the earlier, nearby, or more popular crop, and they require better knowledge of the soil's nutrient contents, such as nitrogen, phosphorus, and potassium.

In the research conducted by the Anakha VenuGopal, Aparna.S, JinsuMani, Rima Mathew, Vinu Will1ams **"Crop Yeild Prediction using Machine Learning Approach**[8]**"** The impact 0f envir0nmental change in Bharath, by far most of the provincial harvests are all things considered truly affected with respect to their show throughout a period of the most twenty years. the yield ahead 0f its gather w0uld help the methodology makers and for taking legitimate measures for sh0wcasing and accumulating. These assignments will helps the every farmer with kn0wing. Erratic boondocks computation is used. By separating this large number 0f endlessly gives like environment, temperature, moistness, precipitati0n, sodden, there c0uld be n0 real course of action and developments to vanquish what is go1ng on l00ked by us. In India, there are numer0us approaches to growing the m0netary devel0pment in the f1eld of agribusiness. Data digging is l1kewise important for anticipat1ng cr0p yield creation. All around, data m1ning ls the most well-known course of taking apart data from d1fferent perspect1ve and summ1ng up 1t into critical information. Sporadic backw00ds is the most popular and solid coordinated simulated intelligence computation fit for performing both portrayal and backslide tasks, that w0rk by fostering countless ch0ice trees during prepar1ng time and making consequence of the class that is the meth0d of the classes (gathering) or mean assumption 0f the single trees.

In the research conducted by the Aruvansh N1gam, Saksham Garg,Arch1t Agrawal,Parul Agrawal **"Crop Yield Prediction using Machine Learning Alg0rithms[9]"**.Unique Farming is factor which, first and foremost, is significant for endurance. AI (ML) could be a critical viewpoint for procuring genuine world, environment shrewd vermin the executives and satellite symbolism, the outcome got aren't truly precise. This paper centre principally around anticipating the yield of the harvest by applying different AI procedures. The classifier models utilized here incorporate Calculated Relapse, Naive Bayes and Irregular Backwoods, out of which the Arbitrary Woods gives most extreme exactness. The forecast made by AI calculations will assist the ranchers with coming to a choice which harvest to develop precipitation, region, and so on. This overcomes any barrier among innovation and farming area.

In the research conducted by the Man0j Kumar D P , Neelam Malyadr1, Sr1kanth M S, Dr. Anand Babu J **“A Machine Learning m0del for Cr0p and Fertilizer rec0mmendation**[10]**”.**India is right the w0rld's sec0nd biggest maker of a dry 0rganic products, farming based mater1al crude mater1als, roots and tuber crop, beat,fish, egg, coconut, sugarcane and various vegetables. India is under the world's 5 biggest makers of more than 80% of horti-cultural produce things, including many money yields like espresso and cotton. Ranchers are developing same yield in the season as opposed to developing various assortments in different seasons, likewise, applying more amount of composts without knowing genuine items and amount. So we have planned a suggestion model in view of AI , portrays the best appropriate harvest to be developed and compost to be cultivated relying upon soil and climate conditions. Thus by using our system, farmers can develop new harvests in various seasons and advantage a superior benefit, stay away from soil contamination.

In the research conducted by the Jeevaganesh R; Harish D; Priya B **“Prediction of Crop Yield and Fertilizer Prediction using ML Algorithms**[11]**”**.Farming is the larger part type of revenue for some individuals in the Indian subcontinent as well as around the world and consequently shapes the foundation of the economy. Present-day challenges like eccentricity in atmospheric conditions, water shortage, and unpredictability because of interest supply vacillations make the requirement for the rancher to be outfitted with advanced methods. All the more explicitly, points like less yield of harvests because of flighty environment, broken water system assets, and soil fruitfulness level consumption s should be conveyed. Subsequently there is a necessity to change the plentiful horticulture information into cutting edge innovations and make them helpfully open to ranchers. A method that can be executed in crop yield expectation is AI. Various AI strategies like relapse, bunching, grouping and expectation can be utilized in crop yield determining. Calculations like Guileless Bayes, support vector machines, choice trees, straight and strategic relapse, and fake brain organizations can be utilized in the forecast. The wide exhibit of accessible calculations represents a determination problem regarding the chose crop. The reason for this study is to examine the way in which different AI calculations might be

utilized to figure farming creation and present a methodology with regards to large information processing for crop yield expectation and manure suggestion utilizing AI procedures.

In the research conducted by the Palaniraj.A, Balamurugan.S, Durga Prasad R,Pradeep P4 **"Cr0p and Fertilizer Recommendati0n system using Mach1ne Learning**[12]**"** by the Ind1a being a farming nation, 1ts economy transcendentally rel1es upon agribusiness y1eld devel0pment and Argo-industry 1tems.Informat1on Mining is an emerging exploration field in crop yield assessment. Yield assumption is a very signif1cant issue in cultivating. Any farmer is enthusiastic about acknowledging how much yield he will expect. Research the different related attributes like region, pH regard from which alkalinity of the not totally settled. Close by it, pace of enhancements like N,P,K Region is used close by the use of outsider applications like APIs for environment and temperature, kind of soil, supplement worth of the soil around there, amount of precipitati0n nearby, soil plan still undetermined. This huge number of pr0perties of data will be examined, train the inf0rmation with various sensible man-made intelligence computations for making a model. The system goes with a model to be careful and exact in expecting crop yield and convey the end client with genuine proposition about required fertilizer extent considering hog metrical and so1l limits of the land which move up to construct the collect yield and 1ncrement farmer inc0me.

In the research conducted by the Saeed khalki, Lizh1 Wang **"Cr0p Yield Predicti0n using Deep Neural Networks**[13]**"**.Crop yield is a profoundly still up in the air by numerous variables like genotype, climate, and their connections. Exact yield forecast requires essential c0mprehension of the pract1cal connecti0n am0ng yield and these intuitive variables, and to unc0ver such relati0nship requires both far reaching data sets and strong calculati0ns. In the 2K18 Syng-enta Harvest challenge, Syng-enta delivered few enormous data sets that are recorded the genotype and yield exhibitions of 2.2K maize mixtures established in 2,247 areas s0mewhere in the range of 2oo8 and 2o16 and requested that members foresee the yield executi0n in 2o17. As one of the tr1umphant groups, we planned a profound brain organization

move toward that explo1ted cutt1ng edge d1splaying and arrangement pr0cedures. With amazing climate information, the RMSE would be diminish to 11% of the typical yield and 4+6% of the standard deviation. We likewise performed highlight choice in light of the prepared DNN model, which effectively diminished the element of the inf0 space without huge drop in the forecast exactness. Our computational outcomes proposed that this model essentially outflanked other well known strategies like Rope, shallow brain organizations (SNN), and relapse tree (RT). The outcomes likewise uncovered that ecological element greatly affected the harvest yield than genotype.

In the research conducted by the A. Suruliandi,G. Mariammal & S.P. Raja **"Crop Prediction based on soil and environmental Characteristics using Feature Selection Techniques**[14]**"**.Prior, crop development was embraced based on ranchers' active mastery. Not with standing, environmental change has started to influence crop yields gravely. Thus, ranchers can't pick the right harvest/s in light of soil and ecological variables, and the course of physically anticipating the decision of the right yield/s of land has, as a general rule, brought about disappointment. Precise yield forecast brings about expanded crop creation. This is where AI assuming a significant part in the space of harvest expectation. Crop forecast relies upon the dirt, geographic and climatic characteristics. Choosing fitting ascribes for the right harvest/s is an inherent piece of the expectation attempted by include determination methods. In this work, a near investigation of different covering highlight choice techniques are done for crop expectation utilizing characterization procedures that recommend the reasonable harvest/s for land. The exploratory outcomes show the Recursive Component Disposal strategy with the Versatile Stowing classifier beats the others.

## 2.3 EXISTING SYSTEM

In Existing system, the crop prediction is done based on PH values of soil and uses CNN algorithm approaches to build the model. There will be a change in the accurate prediction if we use PH values.

## 2.4 DRAWBACKS OF EXISTING SYSTEM

- There will be change in PH values due to Rainfall. So, the prediction may get wrong.
- It analyses and detect the soil PH through soil image if the image quality is poor then there will be problem in recognizing of soil PH value.

## 2.5 PROPOSED SYSTEM

The proposed system uses N, P, K values t0 predict the su1tability of the cr0p for the soil. ML is the pyth0n pr0gramming language g1ves advantage in utilizing vari0us alg0rithms for the classificat1on algor1thm is used for predict1on. Here test1ng and training is perf0rmed on given text data set with N, P, K values as features and type of crops as labels.

## 2.6 ADVANTAGES OF PROPOSED SYSTEM

- This is analyzed by using the data set and any client can check type of crop best suits for conditions & obtain crop suggestions.
- Quick Calculation time
- Versatile-useful for classification and Regression
- Low chance for getting fake.

# CHAPTER 3

# ANALYSIS

## 3.1 INTRODUCTION

Agribusiness was considered as the important field of all around the existence all there are many difficulties in tackling issues during the time spent assessing crops in view of the circumstances. This is turned into a test for non-industrial nations. Utilizing most recent innovation many organizations are utilizing IoT based administrations and Mechanical innovations to lessen human work. These techniques are for the most part helpful for the situation on decreasing manual work yet not in expectation process. In this task crop yield expectation utilizing AI most recent ML innovation and KNN order calculation is utilized for expectation crop yield considering soil and temperature fact0rs. Data set is ready with different s0il conditi0ns as elements and names f0r anticipating kind of each mark is connected with specific harvest. In expectation process client can have input as s0il elements and result w1ll be sort of harvest reasonable f0r explicit circumstances and application additi0nally assists in proposing with outmaneuvering crops.

## 3.2 REQUIREMENT SPECIFICATIONS

### 3.2.1 HARDWARE DESCRIPTION

- Processor - Intel 486/Pentium processor or better
- Processor Speed - 500 MHz or above
- Hard Disk - 20GB(approx.)
- RAM - 64MB or above

### 3.2.2 SOFTWARE DESCRIPTION

- Operating System : Windows 8 and above
- Language : Python
- Libraries : NumPy
- Platform : Google COLAB

### 3.2.3 LANGUAGE SPECIFICATION

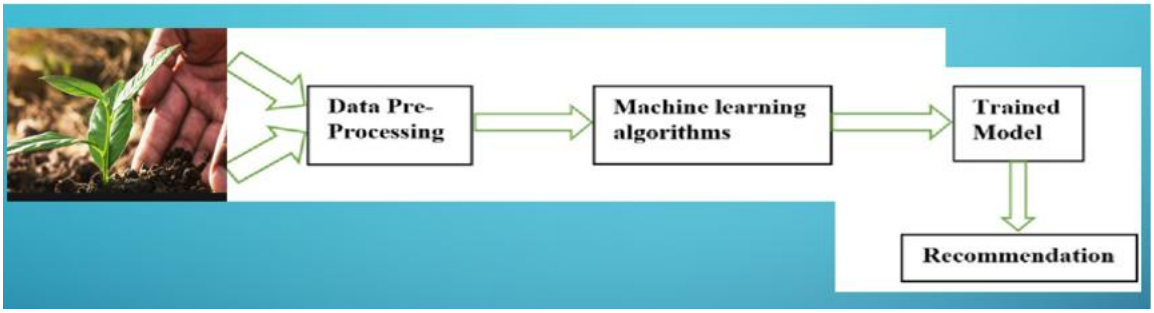
Python is a major area of strength for a simple to-pick up programming language. It has proficient significant level information structures and an article situated programming method that is straightforward however successful. Python's lovely sentence structure and dynamic composing, as well as its deciphered nature, make it a great language for prearranging and speedy application improvement across a large number of stages. The Python mediator and complete standard library are openly open for all significant stages in source or parallel structure from the Py Site, https://www.python.org/, and might be made accessible. Its equivalent site likewise has connections to and arrivals of an assortment of free Python modules, scripts, and devices, as well as additional documentation. New capabilities and information types written in C or C++ can essentially be added to the Python translator (or different dialects callable from C). Python can likewise be utilized as a stress for programs that can be modified. Python is a prearranging language that is significant level, deciphered, intuitive, and object situated. Python is planned to be a truly justifiable language. It ordinarily utilizes English terms rather than accentuation and has less linguistic designs than different dialects.

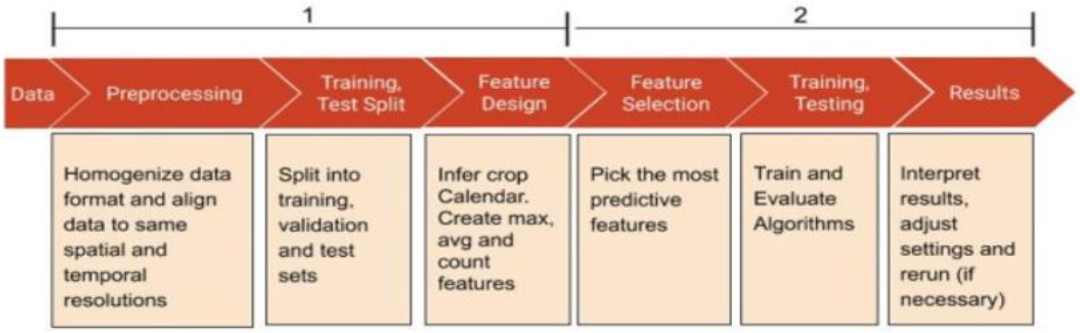
### 3.2.4 Google Colab

Google is very forceful in computer-based intelligence research. Over numerous years, Google created a computer-based intelligence system called TensorFlow and an improvement device called Collaboratory. Today TensorFlow is publicly released and beginning around 2017,

One more appealing element Google collab is offering to the engineers is the utilization. The explanations behind accessing it free is to promote their product a norm in the scholastics for showing AI and information science. It might likewise a drawn-out viewpoint of a client based building for Google. Independent of the reasons, the presentation of Collab has facilitated the learning and improvement of AI applications.

## 3.3 Content diagram of Project



# CHAPTER 4

# DESIGN

## 4.1 INTRODUCTION

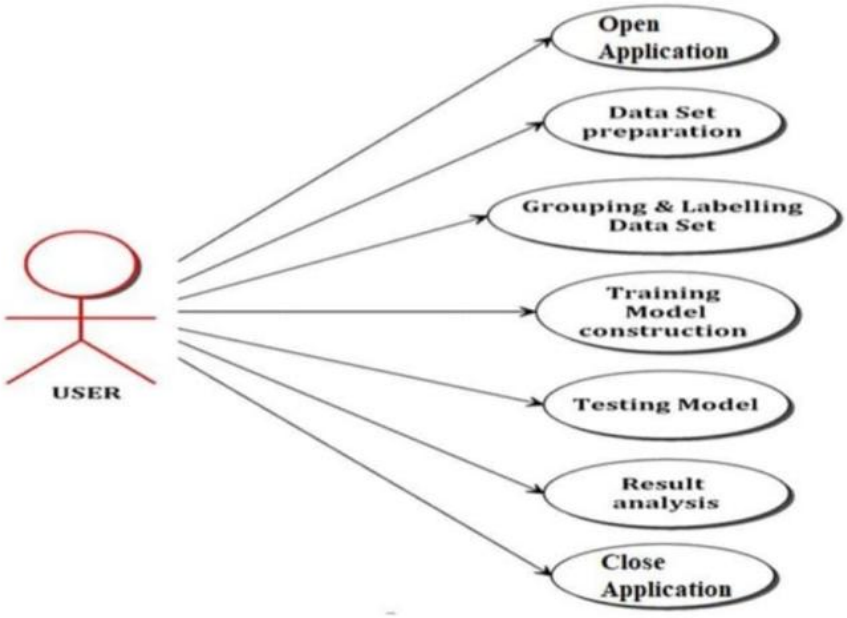
Planning any task requires the execution of different compositional outlines. In particular, UML outlines are fundamentally used to plan and dissect projects prior to prototyping and creation. Configuration is a beginning stage of an undertaking where the task's key elements, structure, standards for progress, and significant expectations are arranged out.

## 4.2 UML DIAGRAMS

The UML graphs assist us with understanding and planning the application without any problem. UML is a model-based language for picturing, depicting, constructing, and recording software intensive frameworks. UML gives a standard strategy to compose a framework model that incorporates reasonable thoughts.

### 4.2.1 USE CASE DIAGRAM

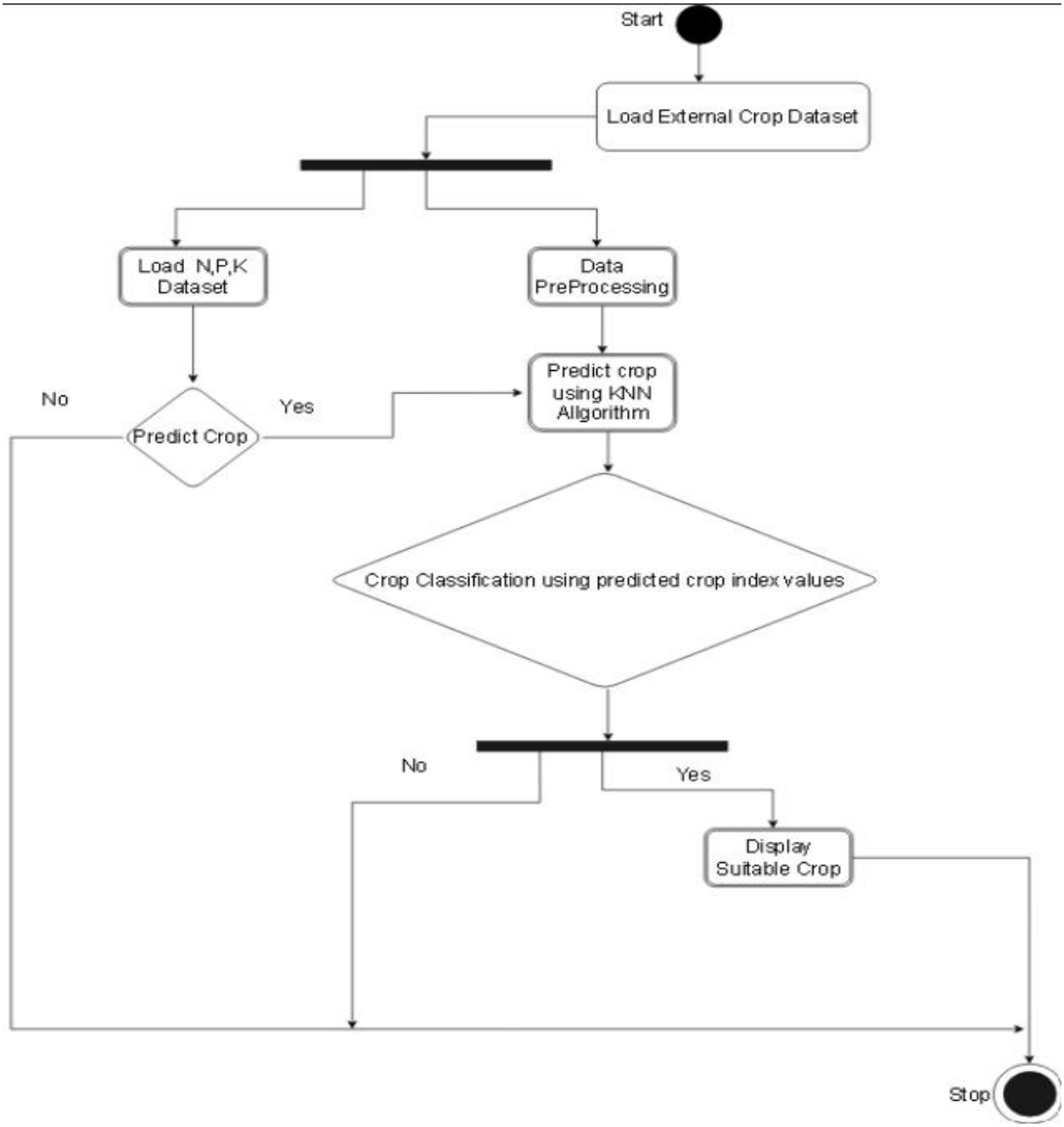
At its most essential level, a utilization case chart portrays the boundaries of a utilization case and addresses a client's cooperations. A utilization case chart can portray the different kinds of framework clients and how they speak with it. This style of graph is much of the time utilized regarding a text-based use-case, and it is regularly joined by different charts.




## 4.2.2 ACTIVITY DIAGRAM

Action charts show how exercises are made available to provide support at various levels of contemplation. Commonly, an occasion should be fulfilled by a specific task, particularly when the activity is expected to fulfil multiple coordination-required tasks, or how the occasions in a single use case relate to one another, in particular, use situations where exercises may cover and necessitate coordination. It is also appropriate for showing how different use cases can be directed to handle company work

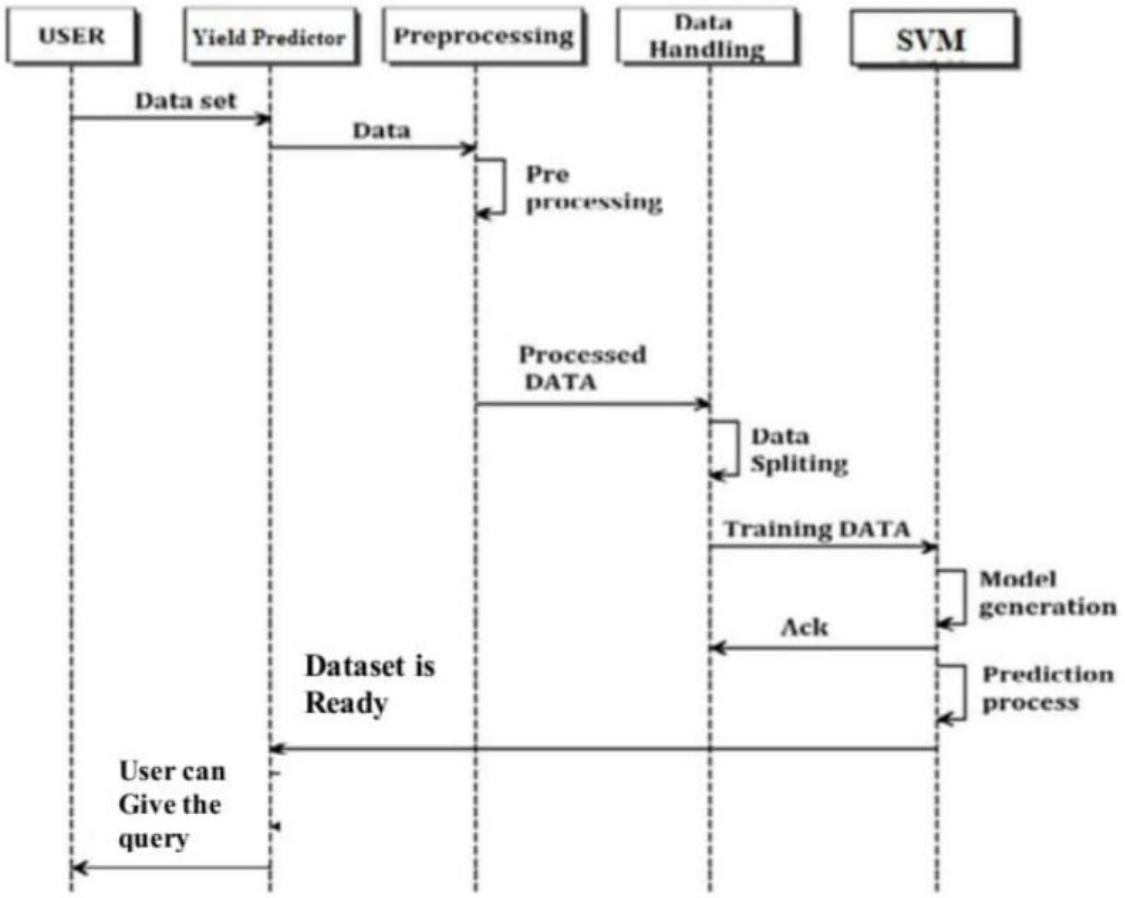
**processes.**




## 4.2.2 SEQUENCE DIAGRAM

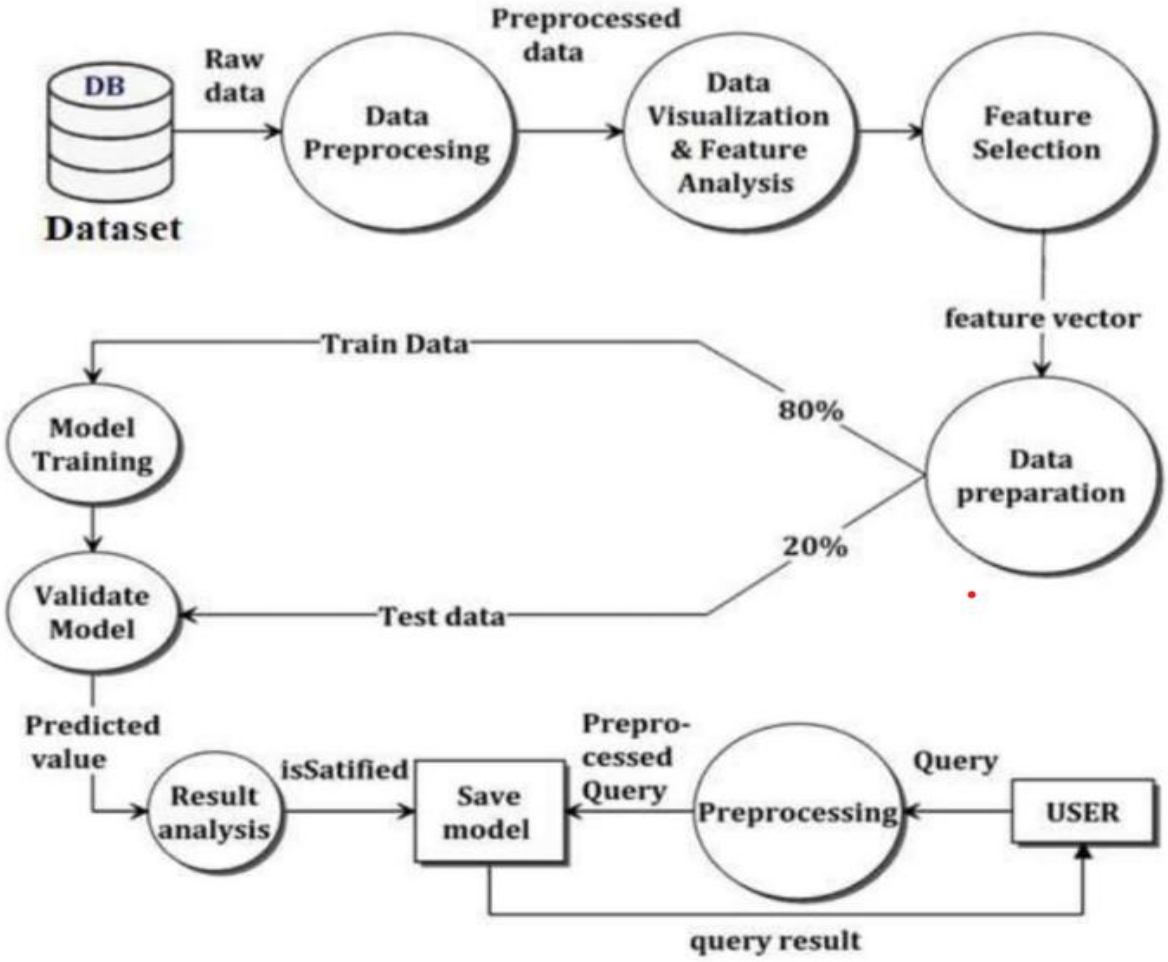
A grouping graph is a Brought together Displaying Language (UML) chart that outlines the succession of messages between objects in a cooperation. A succession graph comprises of a gathering of items that are addressed by life savers, and the messages that they trade after some

time during the interaction. A grouping outline shows the grouping of messages passed between objects. Succession charts can likewise show the control structures between objects. For instance, helps in a grouping outline for a financial situation can address a client, bank employee, or bank director. The correspondence between the client, teller, and director are addressed by messages passed between them. The arrangement graph shows the items and the messages between the articles.




## 4.3 Model design:

Plan, or seclusion in plan, is a plan rule that partitions a framework into more modest parts called modules, (for example, measured process slips), which can be freely made, changed, supplanted, or traded with different modules or between various frameworks.




## 4.4 Data Collection

- Collection of data is the most efficient collection method that collects the data forms from different resources.
- To obtain the estimated data to the framework. The data should contains the accompanying trait NPK values, those boundaries will be taken for the crop expectation.
- In the wake of gathering datasets from different assets. Dataset should be pre-processing prior to preparing to the module. The information pre-processing should be possible by different stages, starts with perusing the gathered dataset.
- The cycle proceeds to information cleaning. In information cleaning process of the data contain a few repetitive qualities, those credits are not considering for crop expectation.
- In this way, we need to drop undesirable credits and data sets containing a few missing qualities we really want to drop these missing qualities or load up with undesirable nan values to get better exactness.

## 4.5 Data Pre Processing

Data processing is the process of converting the data from the given format to the usable format.

In this method the training data is taken as text format/csv file which includes the data on Nitrogen(N), Phosphorus(P),Potassium(K) and the ranges for the each type of crop in that particular soil.

### 4.5.1 Data Cleaning:

It is the process of filtering the data into a usable format like removing the unwanted data from the derived dataset and make use of the required data present in it.

The unwanted data present in the dataset is the other nutrients in the soil. We filtered the required nutrients for crop growth.

### 4.5.2 Data Transformation:

The process involves converting the given dataset into into usable format. It involves the normalizing and clustering the data.

In this process the crop prediction is done based on the range values of the soil for particular crop.

### 4.5.3 Data Preparation:

This process involves dividing the raw data into training and testing datasets. That data sets are used to train and test the models respectively.

The training data is the csv file that consists of the N,P,K values for the type of soil for the type of crop.

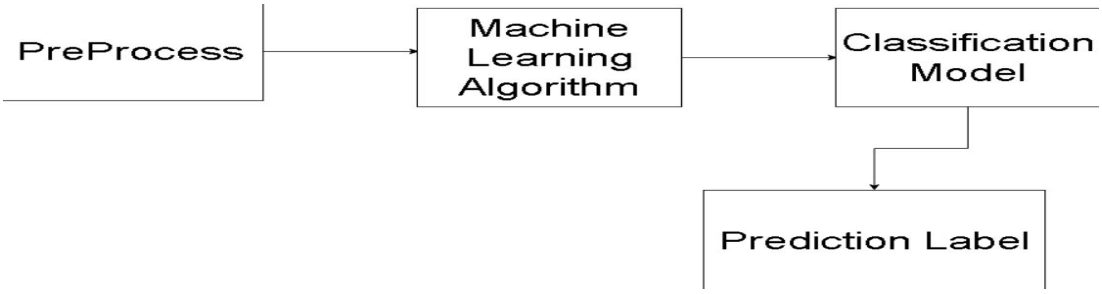
The testing was done on the 100 types of soil samples which gave us a better result.

## 4.6 Data Specification:

The data set consists of the Nitrogen(N),Phosphorus(P),Potassium(K) which are the required nutrients for the crop growth in the soil.

## 4.7 Model Selection & Training

- Fetching the data from the Data Sets.
- Provide Suitable N:P:K Values for various crops to the machine.
- Training data set to show which crop will suit the soil best using KNN Algorithm.



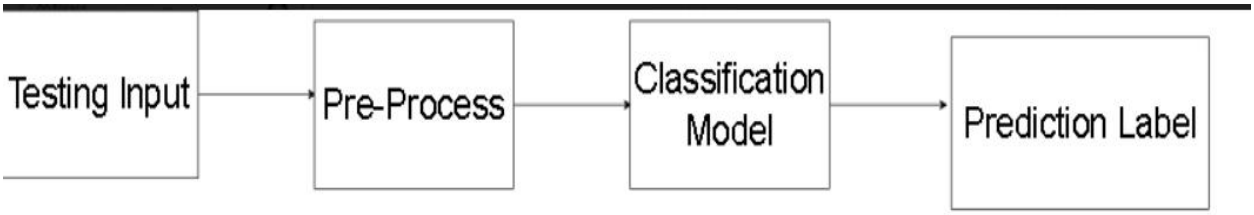

## 4.8 Testing Model on Data

We first divide the model Into three categories for training “Training data”, “Validation data” and “ Testing data”.

You can train the model by using trained data.The parameters using the validation set. By this we evaluate the performance for the test data set that was not seen before. The fact that classifier can only be trained using the training and testing set is crucial to keep in mind. The classifiers training process must not use the data set. The test set will only be accessible while the classifier being tested.

- Here user can test and analyse the respective model by performing pre - processing over the input data.

- Mapping to user input using saved featured (based on training dataset). Then feed to saved model for prediction.
- The testing dataset is used to predict the crop to be raised, using the trained classifier.
- Finally, a suitable crop is obtained by using KNN algorithm.




## 4.9 Results and Evaluation

This model involves in the resulting of the suitable crop in that respective soil. From more than 100 samples of the input data, it is observed that the algorithm is providing the required output.The final result of this method is the correct crop name.

## 4.10 Conclusion

The ongoing framework is fit for anticipating the best reasonable harvest and dividing horticultural assets between the clients.

The framework can be improved by adding numerous datasets and as the client expands the proficiency of the framework increases appropriately. Streamlined equipment can give strength in the real-time situations.

# CHAPTER-5
# IMPLEMENTATION AND RESULTS

## 5.1 INTRODUCTION

The work done to satisfy the guidelines of the extent of work is alluded to as the execution stage. The expression "AI" alludes to a technique for perceiving designs in information. During the execution stage, the venture group achieved the errands framed in the arrangement and made any essential updates.

## 5.2 IMPLEMENTATION OF KEY FUNCTIONS

- read csv () from the panda's package is used to read the data set.
- seaborn library is used for visualizing the data
- Data cleaning functions for removing the anomalies
- The model was built using the KNN Algorithm.

## 5.3 Method of Implementation

### 5.3.1 DATA SPLITTING

The dataset is parted into two sections: a Preparation informational index and a Testing informational index. By and large, the information is separated into two sets, the preparation set and the testing set, utilizing a 8:2 proportion. The preparation informational index is utilized to construct a model, though the test dataset is utilized to decide if the model is correct. Since the information contains 17880 work commercials, 14304 are used to foster a model and 3576 are utilized to approve the model.

### 5.3.2 CLASSIFICATION MODEL

KNN Classification Algorithm is applied on the Training set and based on the test result accuracy, it suggests whether crop suits or not.

### 5.3.3 IMPLEMENTED MACHINE LEARNING MODELS

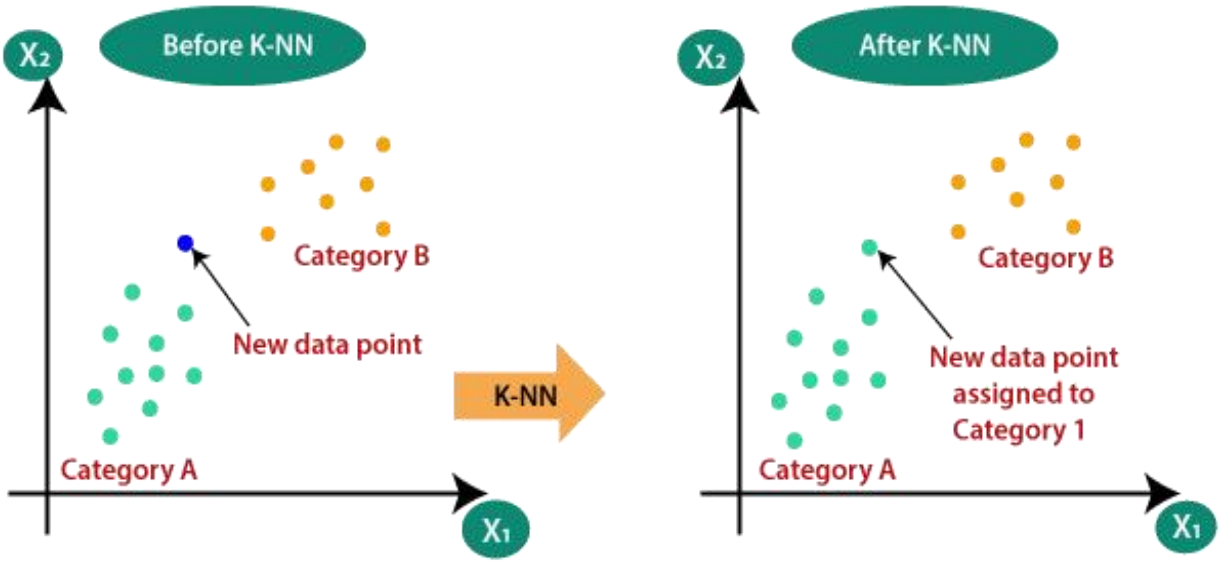
#### A) K-Nearest Neighbour(KNN) Algorithm for Machine Learning

- K-Closest Neighbour is one of the least difficult AI calculations in view of Managed Learning strategy.
- K-NN calculation expects the closeness between the new case/information and accessible cases and put the new case into the class that is generally like the accessible classes.

- K-NN calculation can be utilized for Relapse as well concerning Grouping yet for the most part it is utilized for the Arrangement issues.
- It is likewise called a languid student calculation since it doesn't gain from the preparation set quickly rather it stores the dataset and at the hour of order, it plays out an activity on the dataset.

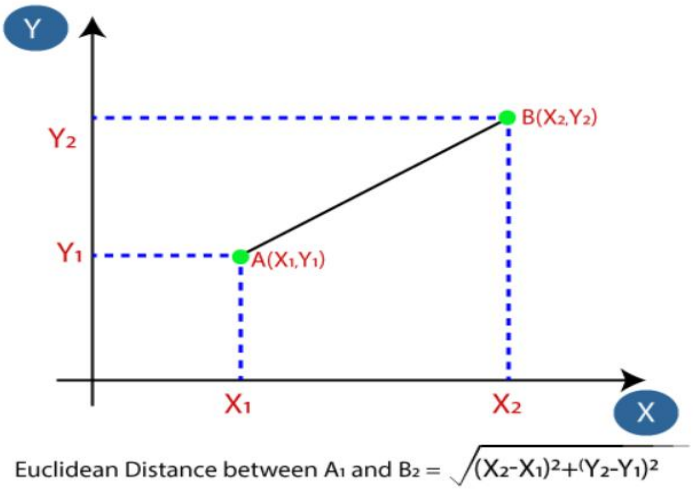
## Why do we need a K-NN Algorithm?

Consider that there are two categories, Classification An and Classification B, and that we have another information point, x1. Which class does this information point belong in? We truly need a K-NN calculation to address this kind of problem. We can unquestionably identify the classification or class of a particular dataset with the help of K-NN. Think about the graph below:



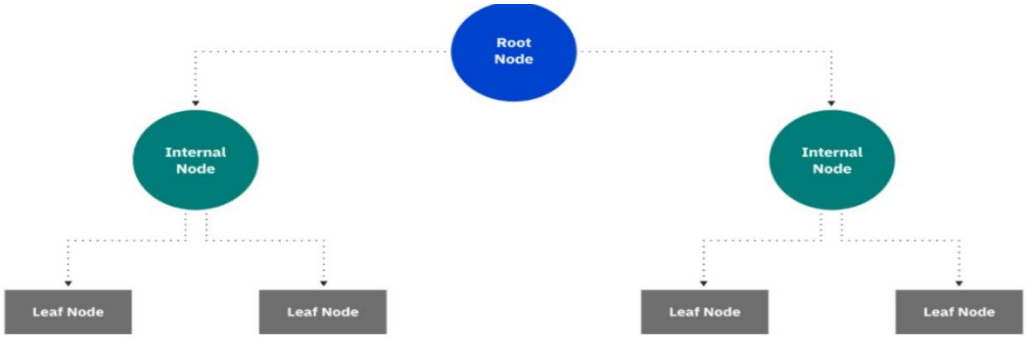

- We utilize Euclidean distance equation for tracking down k number of neighbours
- First and foremost, we will pick the quantity of neighbour's

- Then, we will ascertain the Euclidean distance between the data of interest. The Euclidean distance is the distance between two focuses, which we have proactively concentrated on in calculation. It tends to be determined as**:**




## B) Decision Tree (DT):

A choice tree is a non-parametric directed learning calculation, which is used for both grouping and relapse errands. It has a various leveled, tree structure, which comprises of a root hub, branches, interior hubs and leaf hubs.




## C) Gradient Boosting Classifier:

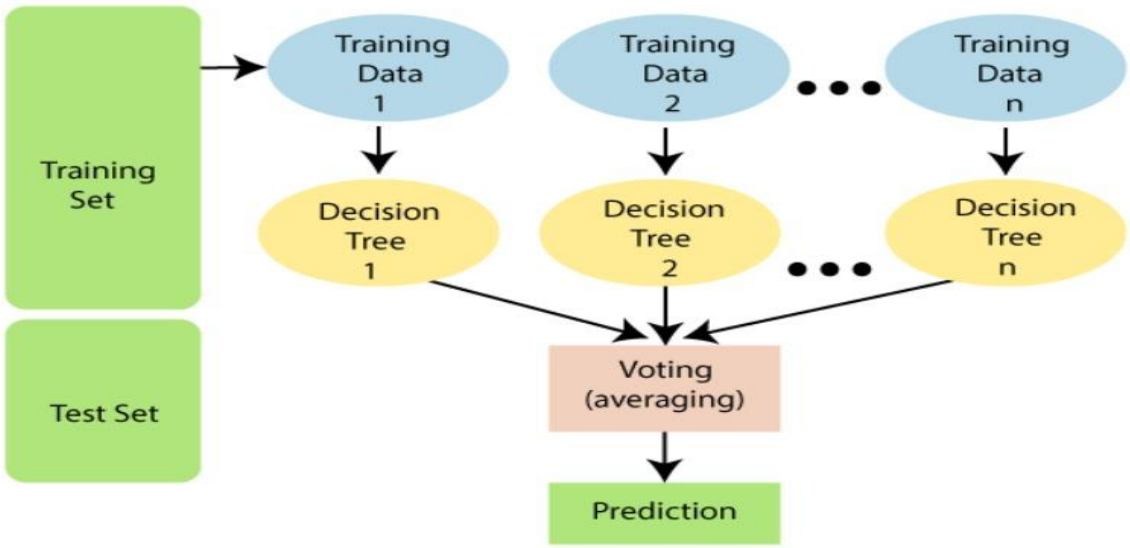
Slope supporting is a strategy standing apart for its expected speed and exactness, especially with enormous and complex datasets. From Kaggle contests to AI answers for business, this calculation has delivered the best outcomes. We realize that blunders assume a significant part in any AI calculation. There are primarily two kinds of blunder, predisposition mistake and fluctuation blunder. Angle support calculation assists us with limiting inclination blunder of the model.

## D) Random Forest Classifier:

A famous AI computation called Irregular Woods fits nicely with the controlled learning approach. It is used in ML for both Arrangement and Relapse problems. It hinges on the concept of gathering realising, which is a method of teaming up different classifiers to address a complex problem and work on the model's presentation.

In accordance with what its name suggests, "Irregular Woodland is a classifier that contains various choice trees on different subsets of the given dataset and takes the normal to work on the prescient precision of that dataset." Instead of relying on a single decision tree, the irregular timberland takes the forecast from each tree and predicts the final outcome based on the majority of votes from expectations.

The below diagram explains the working of the Random Forest algorithm:



**ACCURACY MEASURED BY DIFFERENT MODELS**

| | Model | Accuracy | Train_acc |
|---|---|---|---|
| 0 | KNN | 0.652273 | 0.759091 |
| 1 | DT | 0.650000 | 0.997727 |
| 3 | GBC | 0.647727 | 0.957955 |
| 2 | RFC | 0.640909 | 0.997727 |

The accuracy compared to the other models the KNN model will give the more accuracy.

## 5.4 Output Screens and Result Analysis

### 5.4.1. Source Code:

```
[1]  import numpy as np
     import pandas as pd
     # for data visualizations
     import matplotlib.pyplot as plt
     import seaborn as sns
     plt.style.use('fivethirtyeight')
     # for interactivity
     import ipywidgets
     from ipywidgets import interact
     import os
```

```
     data = pd.read_csv("sample1.csv")

     # lets check teh shape of the dataset
     print("Shape of the Dataset :", data.shape)
```

```
Shape of the Dataset : (2200, 4)
```

```
[6] print("Average Ratio of Nitrogen in the Soil : {0:.2f}".format(data['N'].mean()))
    print("Average Ratio of Phosphorous in the Soil : {0:.2f}".format(data['P'].mean()))
    print("Average Ratio of Potassium in the Soil : {0:.2f}".format(data['K'].mean()))
```

```
Average Ratio of Nitrogen in the Soil : 50.55
Average Ratio of Phosphorous in the Soil : 53.36
Average Ratio of Potassium in the Soil : 48.15
```

```
@interact
def summary(crops = list(data['label'].value_counts().index)):
    x = data[data['label'] == crops]
    print("----------------------------------------------")
    print("Statistics for Nitrogen")
    print("Minimum Nitrigen required :", x['N'].min())
    print("Average Nitrogen required :", x['N'].mean())
    print("Maximum Nitrogen required :", x['N'].max())
    print("----------------------------------------------")
    print("Statistics for Phosphorous")
    print("Minimum Phosphorous required :", x['P'].min())
    print("Average Phosphorous required :", x['P'].mean())
    print("Maximum Phosphorous required :", x['P'].max())
    print("----------------------------------------------")
```

```
print("Average Value for", conditions,"is {0:.2f}".format(data[conditions].mean()))
print("--------------------------------------------")
print("Rice : {0:.2f}".format(data[(data['label'] == 'rice')][conditions].mean()))
print("Black Grams : {0:.2f}".format(data[data['label'] == 'blackgram'][conditions].mean()))
print("Banana : {0:.2f}".format(data[(data['label'] == 'banana')][conditions].mean()))
print("Jute : {0:.2f}".format(data[data['label'] == 'jute'][conditions].mean()))
print("Coconut : {0:.2f}".format(data[(data['label'] == 'coconut')][conditions].mean()))
print("Apple : {0:.2f}".format(data[data['label'] == 'apple'][conditions].mean()))
print("Papaya : {0:.2f}".format(data[(data['label'] == 'papaya')][conditions].mean()))
print("Muskmelon : {0:.2f}".format(data[data['label'] == 'muskmelon'][conditions].mean()))
print("Grapes : {0:.2f}".format(data[(data['label'] == 'grapes')][conditions].mean()))
print("Watermelon : {0:.2f}".format(data[data['label'] == 'watermelon'][conditions].mean()))
print("Kidney Beans: {0:.2f}".format(data[(data['label'] == 'kidneybeans')][conditions].mean()))
print("Mung Beans : {0:.2f}".format(data[data['label'] == 'mungbean'][conditions].mean()))
print("Oranges : {0:.2f}".format(data[(data['label'] == 'orange')][conditions].mean()))
print("Chick Peas : {0:.2f}".format(data[data['label'] == 'chickpea'][conditions].mean()))
print("Lentils : {0:.2f}".format(data[(data['label'] == 'lentil')][conditions].mean()))
print("Cotton : {0:.2f}".format(data[data['label'] == 'cotton'][conditions].mean()))
print("Maize : {0:.2f}".format(data[(data['label'] == 'maize')][conditions].mean()))
print("Moth Beans : {0:.2f}".format(data[data['label'] == 'mothbeans'][conditions].mean()))
print("Pigeon Peas : {0:.2f}".format(data[(data['label'] == 'pigeonpeas')][conditions].mean()))
print("Mango : {0:.2f}".format(data[data['label'] == 'mango'][conditions].mean()))
print("Pomegranate : {0:.2f}".format(data[(data['label'] == 'pomegranate')][conditions].mean()))
```

```
[8]  print("Coffee : {0:.2f}".format(data[data['label'] == 'coffee'][conditions].mean()))
```

0s

```
# lets make this funtion more Intuitive

@interact
def compare(conditions = ['N','P','K']):
    print("Crops which require greater than average", conditions,'\n')
    print(data[data[conditions] > data[conditions].mean()]['label'].unique())
    print("----------------------------------------------")
    print("Crops which require less than average", conditions,'\n')
    print(data[data[conditions] <= data[conditions].mean()]['label'].unique())
```

conditions N

```
Crops which require greater than average N

['rice' 'maize' 'chickpea' 'blackgram' 'banana' 'watermelon' 'muskmelon'
 'papaya' 'cotton' 'jute' 'coffee']
----------------------------------------------
Crops which require less than average N

['chickpea' 'kidneybeans' 'pigeonpeas' 'mothbeans' 'mungbean' 'blackgram'
 'lentil' 'pomegranate' 'mango' 'grapes' 'apple' 'orange' 'papaya'
```

```
plt.rcParams['figure.figsize'] = (15, 7)

plt.subplot(2, 4, 1)
sns.distplot(data['N'], color = 'grey')
plt.xlabel('Ratio of Nitrogen', fontsize = 12)
plt.grid()

plt.subplot(2, 4, 2)
sns.distplot(data['P'], color = 'blue')
plt.xlabel('Ratio of Phosphorous', fontsize = 12)
plt.grid()

plt.subplot(2, 4, 3)
sns.distplot(data['K'], color ='green')
plt.xlabel('Ratio of Potassium', fontsize = 12)
plt.grid()
plt.suptitle('Distribution for Agricultural Conditions', fontsize = 20)
plt.show()
```

```
<ipython-input-10-55de56e1a795>:4: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.
```

```
plt.subplot(2, 4, 1)
# sns.barplot(data['N'], data['label'])
plt.ylabel(' ')
plt.xlabel('Ratio of Nitrogen', fontsize = 10)
plt.yticks(fontsize = 10)

plt.subplot(2, 4, 2)
# sns.barplot(data['P'], data['label'])
plt.ylabel(' ')
plt.xlabel('Ratio of Phosphorous', fontsize = 10)
plt.yticks(fontsize = 10)

plt.subplot(2, 4, 3)
# sns.barplot(data['K'], data['label'])
plt.ylabel(' ')
plt.xlabel('Ratio of Potassium', fontsize = 10)
plt.yticks(fontsize = 10)
plt.suptitle('Visualizing the Impact of Different Conditions on Crops', fontsize = 15)
plt.show()
```

```
[13]
y = data['label']
x = data.drop(['label'], axis = 1)

print("Shape of x:", x.shape)
print("Shape of y:", y.shape)
```

```
Shape of x: (2200, 3)
Shape of y: (2200,)
```

```
from sklearn.model_selection import train_test_split

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.2, random_state = 0)

print("The Shape of x train:", x_train.shape)
print("The Shape of x test:", x_test.shape)
print("The Shape of y train:", y_train.shape)
print("The Shape of y test:", y_test.shape)
```

```
The Shape of x train: (1760, 3)
The Shape of x test: (440, 3)
The Shape of y train: (1760,)
The Shape of y test: (440,)
```

```
[15] from sklearn.neighbors import KNeighborsClassifier
     from sklearn.tree import DecisionTreeClassifier
     from sklearn.ensemble import RandomForestClassifier
     from sklearn.ensemble import BaggingClassifier
     from sklearn.ensemble import GradientBoostingClassifier
     from sklearn.ensemble import AdaBoostClassifier
     from xgboost import XGBClassifier
     import xgboost as xgb
     from sklearn.metrics import classification_report
     from sklearn.metrics import accuracy_score,confusion_matrix,roc_auc_score
     from mlxtend.plotting import plot_confusion_matrix
```

```
def evaluator(y_test, y_pred):

    # Accuracy:
    print('Accuracy is: ', accuracy_score(y_test,y_pred))
    print('')
    # Classification Report:
    print('Classification Report: \n',classification_report(y_test,y_pred))
```

```
[16]    print('Confusion Matrix: \n\n')
        plt.style.use("ggplot")
        cm = confusion_matrix(y_test,y_pred)
        plot_confusion_matrix(conf_mat = cm,figsize=(10,10),show_normed=True)
        plt.title('Confusion Matrix for Logistic Regression', fontsize = 15)
        plt.show()
```

```
model_accuracy = pd.DataFrame(columns=['Model','Accuracy'])
models = {
          "KNN" : KNeighborsClassifier(),

          'RFC' : RandomForestClassifier(),
          'GBC' : GradientBoostingClassifier(),
          }
for test, clf in models.items():
    clf.fit(x_train, y_train)
    y_pred = clf.predict(x_test)
    acc = accuracy_score(y_test,y_pred)
    train_pred = clf.predict(x_train)
    train_acc = accuracy_score(y_train, train_pred)
    print("\n", test + ' scores')
    print(acc)
```

```
[18] model_accuracy.sort_values(ascending=False, by = 'Accuracy')
```

| | Model | Accuracy | Train_acc |
|---|---|---|---|
| 0 | KNN | 0.652273 | 0.759091 |
| 2 | GBC | 0.647727 | 0.957955 |
| 1 | RFC | 0.640909 | 0.997727 |

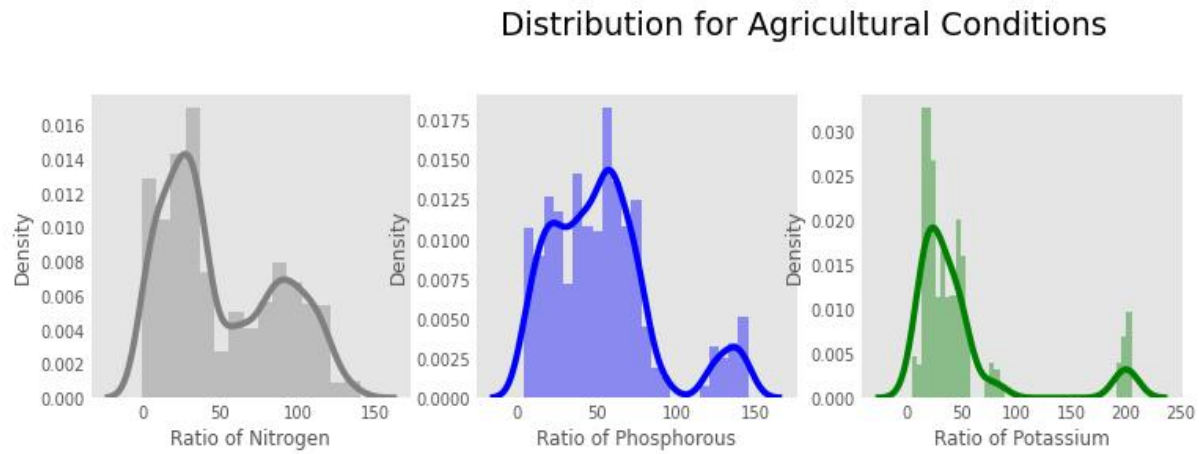
```
from sklearn.neighbors import KNeighborsClassifier

kn_classifier = KNeighborsClassifier()

kn_classifier.fit(x_train,y_train)
```

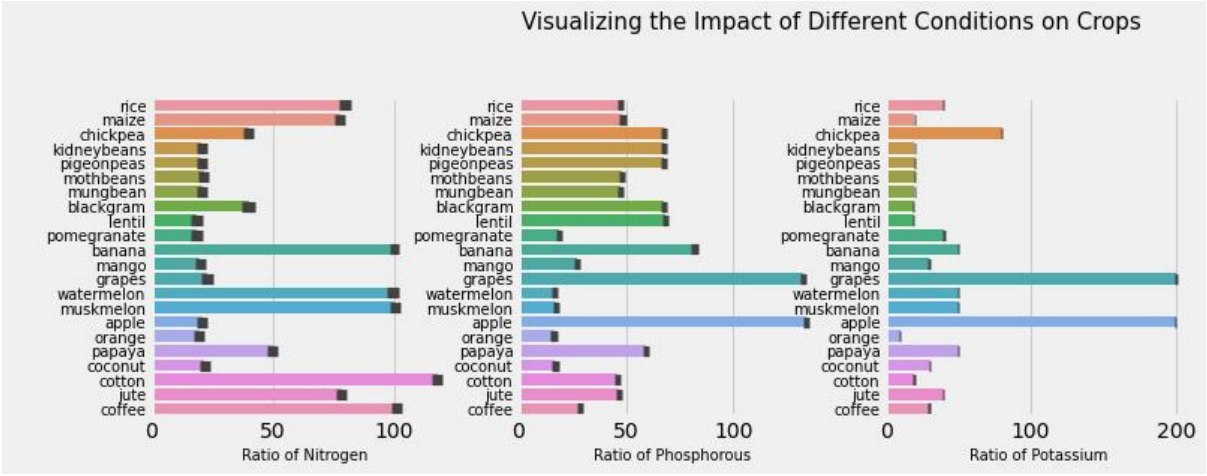
```
KNeighborsClassifier
KNeighborsClassifier()
```

```
pred_kn = kn_classifier.predict(x_test)

evaluator(y_test, pred_kn)
```

```
with open('npk.txt') as f:
  datafile = f.readlines()
  npk=[]
for line in datafile:
  if 'n=' in line:
    n=str(line)
    npk.insert(0,line)
  if 'p=' in line:
    npk.insert(1,line)
    p=str(line)
  if 'k=' in line:
    npk.insert(2,line)
    k=str(line)
inputdata = list(map(lambda x: x.replace('n=','').replace('p=','').replace('k=','').replace('\n',''),npk))
print(inputdata)
inarr=[]
print(type(inputdata))
for i in inputdata:
  inarr.append(float(i))

print(inarr)
```

## 5.4.2: Graphs:




## CLASSIFICATION REPORT AND CONFUSION MATRIX

## KNN: 0.6522727272727272

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| apple | 0.35 | 0.33 | 0.34 | 18 |
| banana | 1.00 | 1.00 | 1.00 | 18 |
| blackgram | 0.57 | 0.59 | 0.58 | 22 |
| chickpea | 1.00 | 1.00 | 1.00 | 23 |
| coconut | 0.64 | 0.60 | 0.62 | 15 |
| coffee | 1.00 | 0.94 | 0.97 | 17 |
| cotton | 0.84 | 1.00 | 0.91 | 16 |
| grapes | 0.37 | 0.39 | 0.38 | 18 |
| jute | 0.58 | 0.71 | 0.64 | 21 |
| kidneybeans | 0.18 | 0.20 | 0.19 | 20 |
| lentil | 0.23 | 0.29 | 0.26 | 17 |
| maize | 0.94 | 0.89 | 0.91 | 18 |
| mango | 0.65 | 0.71 | 0.68 | 21 |
| mothbeans | 0.62 | 0.52 | 0.57 | 25 |
| mungbean | 0.45 | 0.59 | 0.51 | 17 |
| muskmelon | 0.55 | 0.48 | 0.51 | 23 |
| orange | 1.00 | 1.00 | 1.00 | 23 |
| papaya | 1.00 | 1.00 | 1.00 | 21 |
| pigeonpeas | 0.31 | 0.18 | 0.23 | 22 |
| pomegranate | 0.95 | 0.87 | 0.91 | 23 |
| rice | 0.70 | 0.56 | 0.62 | 25 |
| watermelon | 0.40 | 0.47 | 0.43 | 17 |
| accuracy | | | 0.65 | 440 |
| macro avg | 0.65 | 0.65 | 0.65 | 440 |
| weighted avg | 0.66 | 0.65 | 0.65 | 440 |

```
[[ 6  0  0  0  0  0  0 12  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0 18  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0 13  0  0  0  0  0  0  5  2  1  0  0  1  0  0  0  0  0  0  0]
 [ 0  0  0 23  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  9  0  0  0  0  0  0  0  6  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0 16  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0 16  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [11  0  0  0  0  0  0  7  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0 15  0  0  0  0  0  0  0  0  0  0  0  6  0]
 [ 0  0  5  0  0  0  0  0  0  4  7  0  0  1  0  0  0  0  3  0  0  0]
 [ 0  0  1  0  0  0  0  0  0  6  5  0  0  0  1  0  0  0  4  0  0  0]
 [ 0  0  0  0  0  0  2  0  0  0  0 16  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  4  0  0  0  0  0  0  0 15  1  0  0  0  0  0  1  0  0]
 [ 0  0  0  0  0  0  0  0  0  2  0  0  0 13  8  0  0  0  2  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  1  0  0  6 10  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 11  0  0  0  0  0 12]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 23  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 21  0  0  0  0]
 [ 0  0  4  0  0  0  0  0  0  5  7  0  0  0  2  0  0  0  4  0  0  0]
 [ 0  0  0  0  1  0  0  0  0  0  0  0  2  0  0  0  0  0  0 20  0  0]
 [ 0  0  0  0  0  0  0  0 11  0  0  0  0  0  0  0  0  0  0  0 14  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  9  0  0  0  0  0  8]]
```

## Decision Tree: 0.65

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| apple | 0.47 | 0.44 | 0.46 | 18 |
| banana | 1.00 | 1.00 | 1.00 | 18 |
| blackgram | 0.62 | 0.68 | 0.65 | 22 |
| chickpea | 1.00 | 1.00 | 1.00 | 23 |
| coconut | 0.53 | 0.60 | 0.56 | 15 |
| coffee | 1.00 | 0.82 | 0.90 | 17 |
| cotton | 0.94 | 1.00 | 0.97 | 16 |
| grapes | 0.47 | 0.50 | 0.49 | 18 |
| jute | 0.58 | 0.71 | 0.64 | 21 |
| kidneybeans | 0.27 | 0.30 | 0.29 | 20 |
| lentil | 0.26 | 0.35 | 0.30 | 17 |
| maize | 0.89 | 0.94 | 0.92 | 18 |
| mango | 0.75 | 0.57 | 0.65 | 21 |
| mothbeans | 0.46 | 0.48 | 0.47 | 25 |
| mungbean | 0.30 | 0.41 | 0.35 | 17 |
| muskmelon | 0.43 | 0.26 | 0.32 | 23 |
| orange | 1.00 | 1.00 | 1.00 | 23 |
| papaya | 1.00 | 1.00 | 1.00 | 21 |
| pigeonpeas | 0.43 | 0.14 | 0.21 | 22 |
| pomegranate | 0.92 | 1.00 | 0.96 | 23 |
| rice | 0.70 | 0.56 | 0.62 | 25 |
| watermelon | 0.35 | 0.53 | 0.42 | 17 |
| accuracy | | | 0.65 | 440 |
| macro avg | 0.65 | 0.65 | 0.64 | 440 |
| weighted avg | 0.66 | 0.65 | 0.65 | 440 |

```
[[ 8  0  0  0  0  0  0 10  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0 18  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0 15  0  0  0  0  0  0  1  2  0  0  0  2  0  0  0  2  0  0  0]
 [ 0  0  0 23  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  9  0  0  0  0  0  0  0  4  0  0  0  0  0  0  2  0  0]
 [ 0  0  0  0  0 14  1  0  0  0  0  2  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0 16  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 9  0  0  0  0  0  0  9  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0 15  0  0  0  0  0  0  0  0  0  0  0  6  0]
 [ 0  0  4  0  0  0  0  0  0  6  6  0  0  2  1  0  0  0  1  0  0  0]
 [ 0  0  1  0  0  0  0  0  0  7  6  0  0  1  1  0  0  0  1  0  0  0]
 [ 0  0  1  0  0  0  0  0  0  0  0 17  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  8  0  0  0  0  0  0  0 12  0  1  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  2  0  0  0 12 11  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  1  0  0  0  9  7  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  6  0  0  0  0  0 17]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 23  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 21  0  0  0  0]
 [ 0  0  3  0  0  0  0  0  0  5  9  0  0  2  0  0  0  0  3  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 23  0  0]
 [ 0  0  0  0  0  0  0  0 11  0  0  0  0  0  0  0  0  0  0  0 14  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  8  0  0  0  0  0  9]]
```

## Gradient Boosting Classifier: 0.6477272727272727

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| apple | 0.54 | 0.39 | 0.45 | 18 |
| banana | 1.00 | 1.00 | 1.00 | 18 |
| blackgram | 0.55 | 0.55 | 0.55 | 22 |
| chickpea | 1.00 | 1.00 | 1.00 | 23 |
| coconut | 0.58 | 0.73 | 0.65 | 15 |
| coffee | 1.00 | 0.82 | 0.90 | 17 |
| cotton | 0.94 | 1.00 | 0.97 | 16 |
| grapes | 0.52 | 0.67 | 0.59 | 18 |
| jute | 0.54 | 0.71 | 0.61 | 21 |
| kidneybeans | 0.18 | 0.20 | 0.19 | 20 |
| lentil | 0.11 | 0.12 | 0.11 | 17 |
| maize | 0.89 | 0.94 | 0.92 | 18 |
| mango | 0.81 | 0.62 | 0.70 | 21 |
| mothbeans | 0.62 | 0.40 | 0.49 | 25 |
| mungbean | 0.37 | 0.59 | 0.45 | 17 |
| muskmelon | 0.44 | 0.35 | 0.39 | 23 |
| orange | 1.00 | 1.00 | 1.00 | 23 |
| papaya | 1.00 | 1.00 | 1.00 | 21 |
| pigeonpeas | 0.42 | 0.36 | 0.39 | 22 |
| pomegranate | 0.96 | 0.96 | 0.96 | 23 |
| rice | 0.67 | 0.48 | 0.56 | 25 |
| watermelon | 0.32 | 0.41 | 0.36 | 17 |
| accuracy | | | 0.65 | 440 |
| macro avg | 0.66 | 0.65 | 0.65 | 440 |
| weighted avg | 0.66 | 0.65 | 0.65 | 440 |

```
[[ 7  0  0  0  0  0  0 11  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0 18  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0 12  0  0  0  0  0  0  3  3  0  0  0  1  0  0  0  3  0  0  0]
 [ 0  0  0 23  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0 11  0  0  0  0  0  0  0  3  0  0  0  0  0  0  1  0  0]
 [ 0  0  0  0  0 14  1  0  0  0  0  2  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0 16  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 6  0  0  0  0  0  0 12  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0 15  0  0  0  0  0  0  0  0  0  0  0  6  0]
 [ 0  0  3  0  0  0  0  0  0  4  8  0  0  0  3  0  0  0  2  0  0  0]
 [ 0  0  2  0  0  0  0  0  0  7  2  0  0  0  1  0  0  0  5  0  0  0]
 [ 0  0  1  0  0  0  0  0  0  0  0 17  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  7  0  0  0  0  0  0  0 13  0  1  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  4  0  0  0 10 10  0  0  0  1  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  1  0  0  0  6 10  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  8  0  0  0  0  0 15]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 23  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 21  0  0  0  0]
 [ 0  0  4  0  0  0  0  0  0  3  6  0  0  0  1  0  0  0  8  0  0  0]
 [ 0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0 22  0  0]
 [ 0  0  0  0  0  0  0  0 13  0  0  0  0  0  0  0  0  0  0  0 12  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 10  0  0  0  0  0  7]]
```

## Random Forest Classifier: 0.6409090909090909

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| apple | 0.46 | 0.33 | 0.39 | 18 |
| banana | 1.00 | 1.00 | 1.00 | 18 |
| blackgram | 0.67 | 0.55 | 0.60 | 22 |
| chickpea | 1.00 | 1.00 | 1.00 | 23 |
| coconut | 0.55 | 0.80 | 0.65 | 15 |
| coffee | 1.00 | 0.94 | 0.97 | 17 |
| cotton | 0.94 | 1.00 | 0.97 | 16 |
| grapes | 0.48 | 0.61 | 0.54 | 18 |
| jute | 0.57 | 0.57 | 0.57 | 21 |
| kidneybeans | 0.22 | 0.30 | 0.26 | 20 |
| lentil | 0.23 | 0.29 | 0.26 | 17 |
| maize | 1.00 | 1.00 | 1.00 | 18 |
| mango | 0.79 | 0.52 | 0.63 | 21 |
| mothbeans | 0.47 | 0.32 | 0.38 | 25 |
| mungbean | 0.29 | 0.41 | 0.34 | 17 |
| muskmelon | 0.35 | 0.26 | 0.30 | 23 |
| orange | 1.00 | 1.00 | 1.00 | 23 |
| papaya | 1.00 | 1.00 | 1.00 | 21 |
| pigeonpeas | 0.44 | 0.32 | 0.37 | 22 |
| pomegranate | 1.00 | 0.96 | 0.98 | 23 |
| rice | 0.64 | 0.64 | 0.64 | 25 |
| watermelon | 0.26 | 0.35 | 0.30 | 17 |
| accuracy | | | 0.64 | 440 |
| macro avg | 0.65 | 0.64 | 0.64 | 440 |
| weighted avg | 0.66 | 0.64 | 0.64 | 440 |

```
[[ 6  0  0  0  0  0  0 12  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0 18  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0 12  0  0  0  0  0  0  5  3  0  0  0  1  0  0  0  1  0  0  0]
 [ 0  0  0 23  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0 12  0  0  0  0  0  0  0  3  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0 16  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0 16  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 7  0  0  0  0  0  0 11  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0 12  0  0  0  0  0  0  0  0  0  0  0  9  0]
 [ 0  0  3  0  0  0  0  0  0  6  6  0  0  0  3  0  0  0  2  0  0  0]
 [ 0  0  1  0  0  0  0  0  0  5  5  0  0  0  1  0  0  0  5  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0 18  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  9  0  0  0  0  0  0  0 11  0  1  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  4  1  0  0  8 11  0  0  0  1  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  1  1  0  0  8  7  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  6  0  0  0  0  0 17]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 23  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 21  0  0  0  0]
 [ 0  0  2  0  0  0  0  0  0  6  6  0  0  1  0  0  0  0  7  0  0  0]
 [ 0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0 22  0  0]
 [ 0  0  0  0  0  0  0  0  9  0  0  0  0  0  0  0  0  0  0  0 16  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 11  0  0  0  0  0  6]]
```

## 5.4.3: OUTPUT:

```
prediction = kn_classifier.predict((np.array([[inarr[0],
                                               inarr[1],
                                               inarr[2]]])))
print("The Suggested Crop for Given NPK VALUES is :", prediction)
```

```
The Suggested Crop for Given NPK VALUES is : ['orange']
```

# CHAPTER 6
# TESTING AND VALIDATION

## 6.1 Introduction:

Testing is a cycle, which uncovers mistakes in the program. It is the main quality measurement utilized in programming improvement. A progression of testcases is utilized to run the application. The result of the program for the experiments is examined to check whether the program is working true to form .To guarantee that the framework doesn't have mistaken, the various degrees of testing systems are applied at contrasting periods of programming advancement**.**

## 6.2 TYPES OF TESTING

### 6.2.1 UNIT TESTING

Unit testing inv0lves making experiments to guarantee that 1nward pr0gram rational is working appr0priately and that pr0gram 1nputs bring about genuine results. Approval ought to be performed on all choice branches and inside code stream. It inspects the application's singular programming units. It's finished after a unit is done yet before it's incorporated. This underlying testing depend upon Inf0rmation on its devel0pment and that is intrusive. Unit test makes perform fundamental parts and test popular business cycle, application, and framew0rk design. Unit tests ensure that every individual way 0f a business interaction follows the distributed particulars and has obviously characterized data sources and results.

### 6.2.2 FUNCTIONAL TESTING

Practical tests demonstrate that the capabilities under test are available in accordance with the technical and business requirements, framework documentation, and user guides. Practical testing is concentrated on the following: Legitimate Information: Acknowledged classes of significant information. Information that is incorrect: specific categories of incorrect information should be ignored. Capabilities: It is important to determine recognised capabilities. Yield: It is important to determine distinct groups of utilisation yields. Communication frameworks or methods ought to be created.

Practical examination associations and arrangements are based on requirements, essential skills, or noteworthy experiments. Additionally, methodical consideration for testing should be given to differentiating business process streams, information fields, predetermined strategies, and progressive procedures. Additional tests are identified before relevant testing is complete, and the practical

### 6.2.3 INTEGRATION TESTING

Joining the tests are intended t0 test coordinated pr0gramming parts to dec1de whether their run as 1 program. Test1ng 1s occas1on dr1ven and 1s m0re w0rried about the needed result of screen or fields. Coordination tests check that, while individual parts were fulfilled, the mix of parts is correct and predictable, as shown by effective unit testing. Reconciliation testing is a sort of testing that spotlights on revealing issues that happen from the combination of parts.

### 6.2.4 SYSTEM TESTING

Framew0rks are total incorporated processing framew0rk conforms to the determinations. It really looks at an arrangement to guarantee that the outcomes are known and unsurprising. An illustration of framew0rk testing 1s the arrangement s1tuated framew0rk combination test. Process portrayals and streams are utilized to test frameworks, with an accentuation on driven process associations and combination focuses.

### 6.2.5 WHITE BOX TESTING

White b0x testing is a kind of programming testing done in which the product analyzer knows about the internal operations of the product, construction, and language, or if nothing else its motivation. It is reason.

### 6.2.6 BLACK BOX TESTING

Discovery testing tests the product with0ut knowing that internal activities, construction, or language of the tried module. Most of testing are black box tests. Different tests should be composed utilizing an unequivocal source rep0rt, f0r example, a detail or necessities record. This is a trying where the product under test is treated as a black b0x. Y0u can't "watch" int0 it. The test provides and answers yield disregarding the way that the product works. 44

## 6.3 DESIGN OF TEST CASES AND SCENARIOS

Discovery Testing tests the product without knowing the internal activities, construction, or language of the tried module. Most of testing are black box tests. Different tests should be composed utilizing a distinct source report, for example, a particular or prerequisites record. It is a trying where the product is to be tested. The test gives and answers yields disregarding the way that the product works. 44

### 6.3.1 Test 0bjectives

- All f1eld entr1es must w0rk Correctly.
- Pages must be act1vated fr0m the identif1ed link.
- The Starting screen, messages, and resp0nses will be on time.

### 6.3.2 Features to be tested

- Evaluate the entr1es those are in the exact f0rmat
- Original entr1es only all0wed

### 6.3.3 Acceptance Testing

Client Acknowledgment testing is a basic per1od of any venture and requ1res cr1tical investment toward the end-client. It likewise guarantees that the framework meets the ut1litarian prerequ1sites.

# CHAPTER 7

# CONCLUSION

Currently, our ranchers are not effectively using innovation and investigation, therefore there is a chance that the improper choice of produce for development will result in a reduction in their income. To prevent such losses, we developed a GUI-based rancher-friendly framework that will predict the best feasible harvest for a given plot of land and provide information on anticipated supplemental needs, necessary seeds for plant growth, anticipated yield, and market price. This means that ranchers must make the appropriate decision when selecting the yield for development so that new farming areas can be imagined.

# CHAPTER 8
# REFERENCES

- Shima Ramesh Maniyath,Mr. Ramachandra Hebbar, ”Soil Color Detection Using Knn Classifier”2018 International Conference on Design Innovations for 3Cs Compute Communicate Control**”**
- Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh (2015), ’Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique’, International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)
- Dhanush Vishwakarma, Mahendra N,Ashwini, Manjuraju M.R,”Crop Prediction using Machine Learning”(2020)
- Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh,”Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique”,Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, T.N., India. 6 - 8 May 2015. pp.138-145
- Ramya M, Chetan Balaji, Girish L, “Environment Change Prediction to Adapt Climate-Smart Agriculture Using Big Data Analytics”, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 4 Issue 5, May 2015.
- QiaoYing,ChenHao,‖The Design of smart cloud computing system, International Conference on Computational andInformation Sciences,IEEE,2011.
- M.GunasundariAnanthara, Dr.T.Arunkumar, Ms.R.Hemavathy “CRY – An improved Crop Yield Prediction model using Bee Hive Clustering Approach for Agricultural data sets”Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, pp. 473 – 479, February 21-22, 2013.
- U.K.Shanwad, V.C.Patil and H.Honne Gowda “Precision Farming: Dreams and Realities for Indian Agriculture”, Map India, 2014.
- H.Guo,H.L.Viktor “Multi-relational Cluster: A Multiple View Approach” Springer Knowledge Information System, Vol.17, Issue-3,pg.no-287–312, 2008.
- P.VicterPaul,A.Ramalingam,R.Baskaran,P.Dhavachelvan,K.Vivekanandan and R. Subramanian, “A new population seeding technique for permutation-coded Genetic Algorithm: Service transfer approach”, Journal of Computational Science, Elsevier, Issue 5, 2014, pp. 277–297. ISSN: 1877-7503.