**Name of the Project**

**BEST CROP PREDICTION**

**PROJECT REPORT**

**Submitted BY**

**TEAM NAME**

BEAT THE CURVE

# TEAM MEMBER NAMES

M.BHAVANA(16AK1A0509)

S.RAKESH(17AK1A0587)

N.SOWMYA SREE(17AK1A04B5)[ECE]

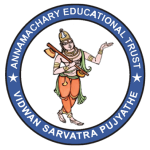
L.LIKHITHA(17AK1A0551)

SK.REEMA(17AK1A0214)[EEE]

***In partial fulfilment for the award of the Certificate***

**of**

# SUMMER INTERNSHIP PROGRAM



## Department of Computer Science and Engineering

**Annamacharya Institute of Technology and Sciences Venkatapuram Village , Renigunta Mandal , Tirupati , Andhra Pradesh 517520 July 2019.**

**BONAFIDE CERTIFICATE**

This is to certify that the project entitled ”**PROJECT TITLE**” submitted by **Team Members Names** in partial fulfilment for the requirements for the award of internship certification in technologies of Machine learning and Deep learning is an authentic work carried out by them under my supervision and guidance.

To the best of my knowledge, the matter embodied in the project report has not been submitted to any other University/Institute for the award of any Degree or Diploma.

**Signature of Supervisor Signature of Head of the Department**

Akshay Kumar Kothuri, M.Tech Mrs.B.Rupa Devi M.Tech,Ph.D.,

AI & IoT Developer Assistant Professor & HOD,

SmartBridge Educational Services PVT LTD Department of CSE,

Hyderabad – 500096 AITS,Tirupati.

BEST CROP PREDICTION

CONTENTS:

* Introduction
* Review of literature
* Data collection
* Methadology
* Findings and suggestions
* Conclusion

1.Abstract:

Best Crop Prediction is a very important issue in agriculture. Any farmer is interested in knowing how much Crop yield he is about to expect. In the past, crop yield prediction was performed by considering farmer's experience on particular field and crop. The yield prediction is a major issue that remains to be solved based on available data. Different Crop Yield Prediction techniques are used and evaluated in agriculture for estimating the future year's crop production. This research proposes and implements a system to predict crop yield from previous data. This is achieved by applying association rule Crop Yielding on agriculture data. This research focuses on creation of a prediction model which may be used to future prediction of crop yield. This paper presents a brief analysis of crop yield prediction based on association rules for the selected region in India. The experimental results shows that the proposed work efficiently predict the crop yield production.

1.1.Introduction:

Agriculture is the backbone of our country and the economic growth of our nation.This paper helps us to predict the best crop thereby improving the quality and profitability of the agriculture sector by processing the datasets.The best prediction concept will help the farmers to choose whether the particular crop is suitable for that specific soil.This prediction can be carried out by using the suitable algorithm where high accuracy and speed can be achieved. In this work KNN model is used to predict best crop.

1.2.System Requirements:

Here we use Basic python language.Python is a powerful multi-purpose programming language created by Guido Van Rossum.It has simple easy-to-use syntax,making it the perfect language for someone trying to learn computer programming for the first time.Python programming language uses a simple object-oriented programming approach and very efficient high-level data structures.Python is often described as a “batteries included” language due to its comprehensive standard library.Python was conceived in the late 1980’s as a successor to the ABC language.Python 2.0,released 2000,introduced features like list comprehensions and a garbage collection system capable of collecting reference cycles.Python interpreters are available for many operating systems.In this work KNN model is used to predict best crop

1.3.Libraries:

* Numpy
* Pandas
* matplotlib

1.4.OBJECTIVES OF RESEARCH:

Productivity analysis of different crops with respect to other similar areas or historical data

Identification of parameters for productivity

Agricultural zoning

A Sustainable development plan

1.5.INDUSTRY PROFILE:

Companies in the industry grow crops such as arhar,cotton ,groundnut ,gram,grains,oilseeds,tobacco,paddy. Major industries include Chiquita Brands international,dole food company ,fresh del monte produce and total produce.sugarcane,corn,rice and wheat are the most highly produced crops by volume. Brazil , the US,India and China are the top crop producers

2. REVIEW OF LITERATURE:

2.1.Crop Selection and Crop Yield Prediction:

To maximize the crop yield, selection of the appropriate crop that will be sown plays a vital role. It depends on various factors like the type of soil and its composition, climate, geography of the region, crop yield, market prices etc. Techniques like Artificial neural networks, K-nearest neighbors and Decision Trees have carved a niche for themselves in the context of crop selection which is based on various factors. A crop selection method called CSM has been proposed which helps in crop selection based on its yield prediction and other factors.

2.2. Weather Forecasting:

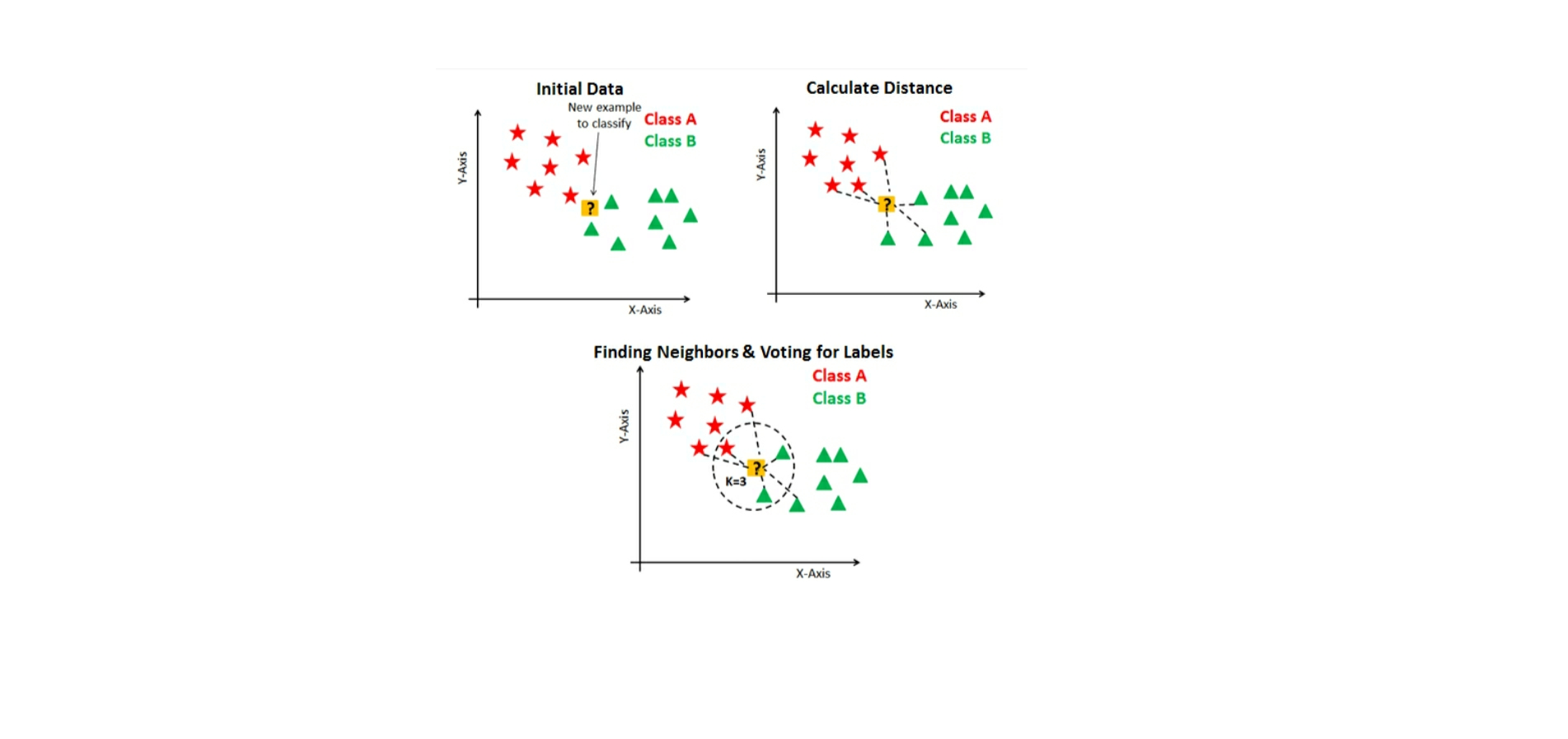
Indian agriculture mainly relies on seasonal rains for irrigation. Therefore, an accurate forecast of weather can reduce the enormous toil faced by farmers in India including crop selection, watering and harvesting. As the farmers have poor access to the Internet as a result of digital-divide, they have to rely on the little information available regarding weather reports. Up-to-date as well as accurate weather information is still not available as the weather changes dynamically over time. Researchers have been working on improving the accuracy of weather predictions by using a variety of algorithms. These algorithms have shown better results over the conventional algorithms.

2.3. Smart Irrigation System:

Farming sector consumes a huge portion of water in India.

2.4 Related Work:

Agricultural management needs simple and accurate estimation techniques to predict rice yields in the planning process. The necessity of the present study were to: identify whether the K-Nearest Neighbors(KNN) algorithm is a type of supervised machine learning algorithms.KNN is extremely easy to implement in its most basic form,and yet performs quite complex classification tasks.It is a lazy learning algorithm,since it doesn’t have specialised training phase



3. CROP YIELD PREDICTION:

Crop yield prediction is widely applied to agricultural issues. Crop prediction is used to analyze large data sets and establish useful classifications and patters in the data sets. The overall goal of the Crop prediction is a process to extract the information from a data set and transform it into understandable structure for further use. This paper analyzes the crop yield production based on available data. The Crop prediction technique was used to predict the crop yield for maximizing the crop productivity.

3.1.DATA COLLECTION:

To start with any Crop prediction problem, it is first necessary to bring all the data together.



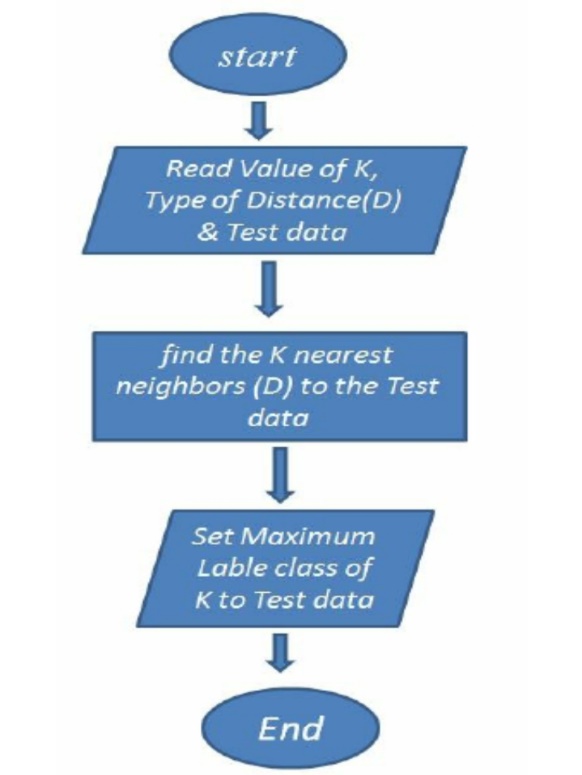
4.Methodology:

The proposed methodology contains two phases: Training Phase and Test Phase. In the training phase the data was collected and preprocessed. The pre-processed data was clustered using k-means clustering algorithm. The association rule crop prediction process will apply on clustered data to find the rules. The training phase ends with number of generated rules. In the testing phase, the yield value is predicted based on the generated rules. The work starts with preprocessing step. In this step the collected data was pre-processed. In the preprocessing, some data was removed from the data set. Some of the area was not suitable for crop production. So that data will be removed.

4.1. EXPLORATORY DATA ANALYSIS:

Exploratory data analysis (EDA) is a strategy of data analysis that emphasizes maintaining an open mind to alternative possibilities .EDA is a philosophy or an attitude about how data analysis should be carried out ,rather than being a fixed set of techniques

4.1. Figures and tables :



4.2.STATISTICAL TECHNIQUES AND VISUALIZATION:

The visualization process consists of three levels of filtering the viability of crops.The first phase would be elimination of non-cultivable crops based on soil types present in the region selected.The second phase involves filtering of crops according to temperature statistics.The third phase involves water availability(considering rainfall statistics).

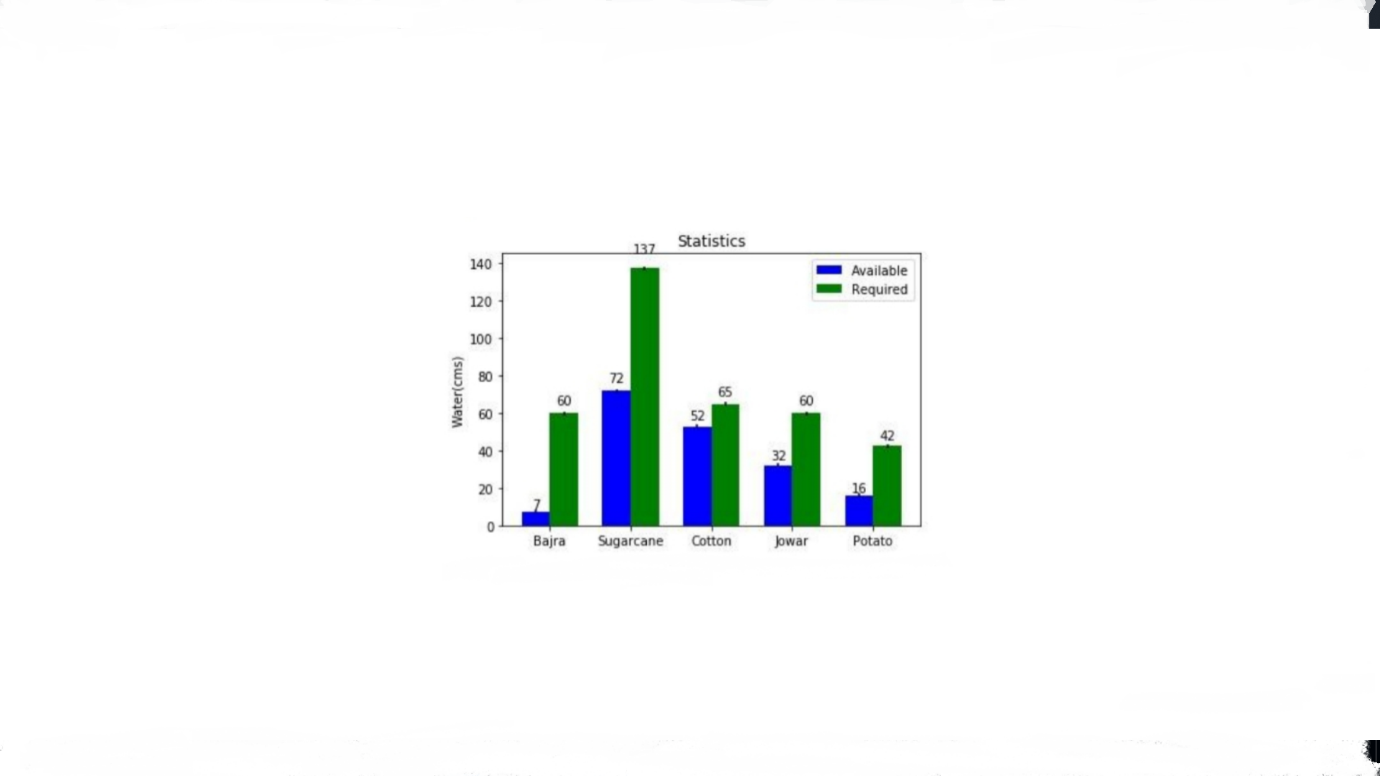
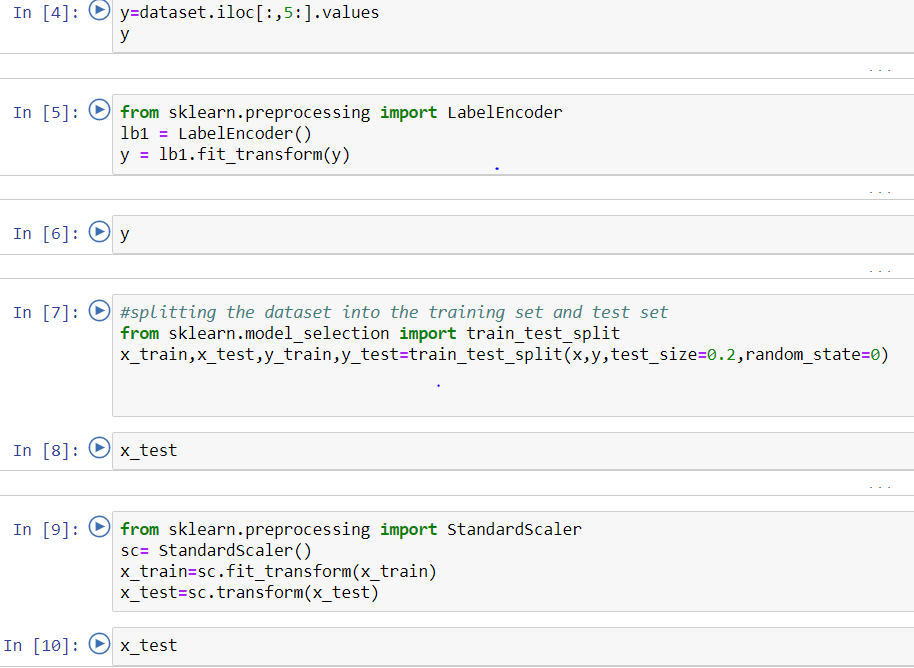


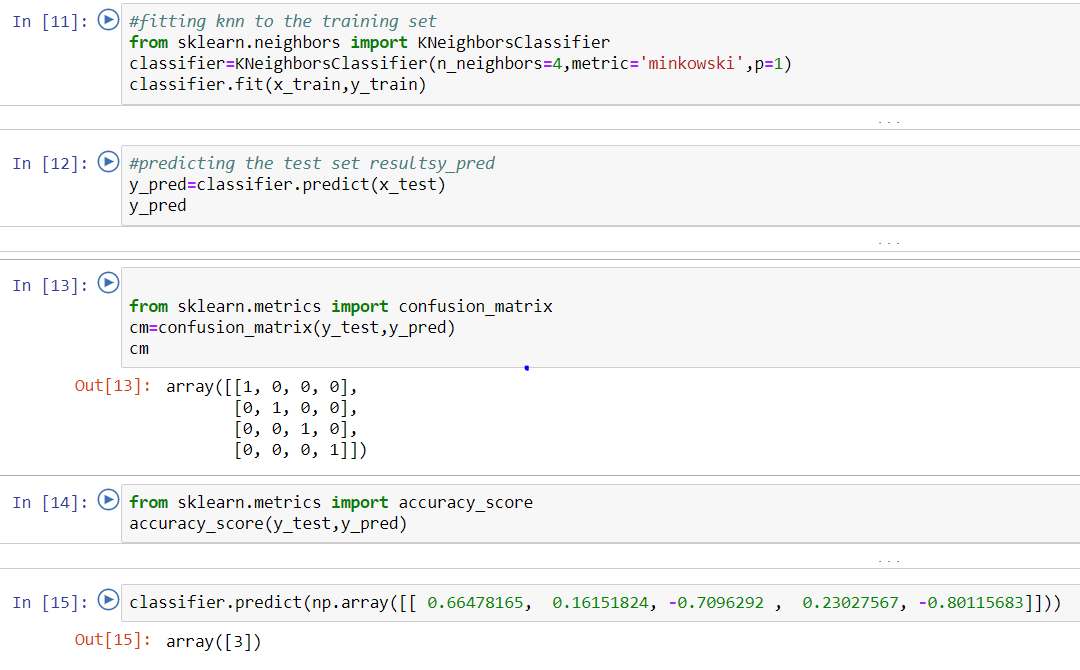
fig: statistical and visualization data analysis

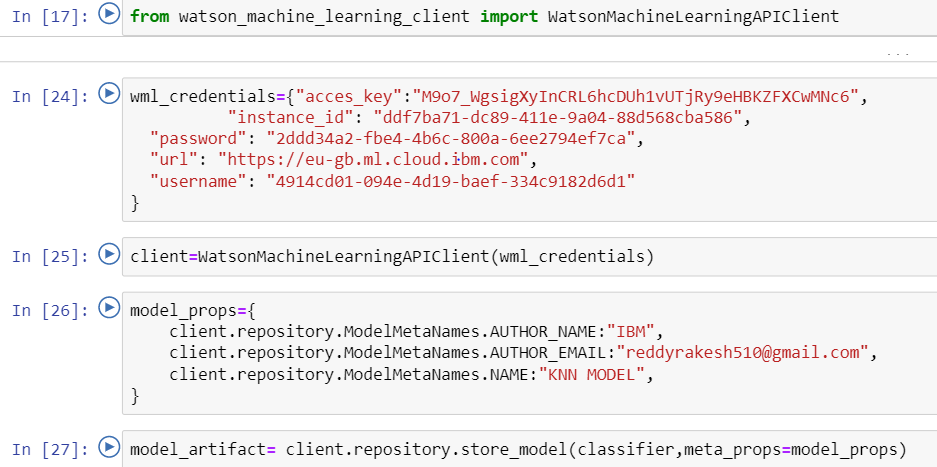
4.3.Data Conversion :

The clustered data was converted into 0’s and 1’s. The data conversion process is mainly used for association rule crop yield production step. The data variables Area, Tanks, Bore Wells, Open Wells, Production and Yield are converted into Low, Medium and High. For each clustering result the data is converted into area Low, area Medium, area High, tank Low, tank Medium, tank High, borewell Low, borewell Medium, borewell High, open well Low, open well Medium, open well High, production Low, production Medium, production High, yield Low, yield Medium, yield High.



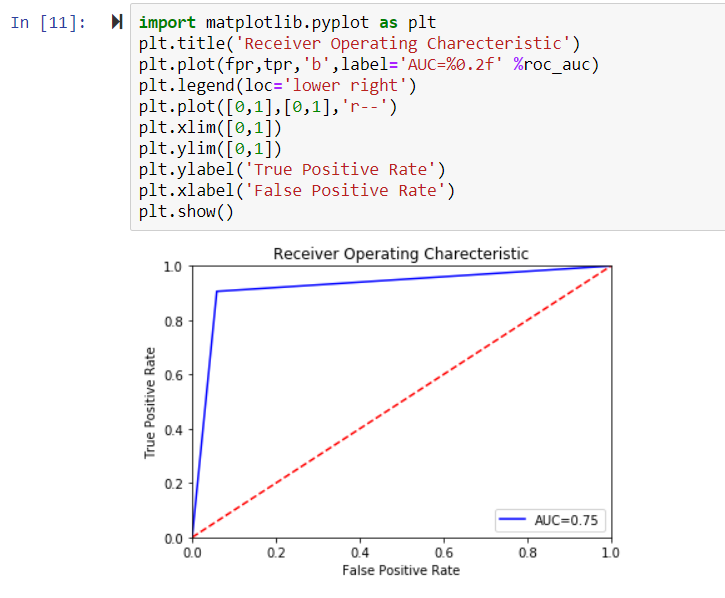
4.4.DATA MODELLING AND VISUALIZATION:



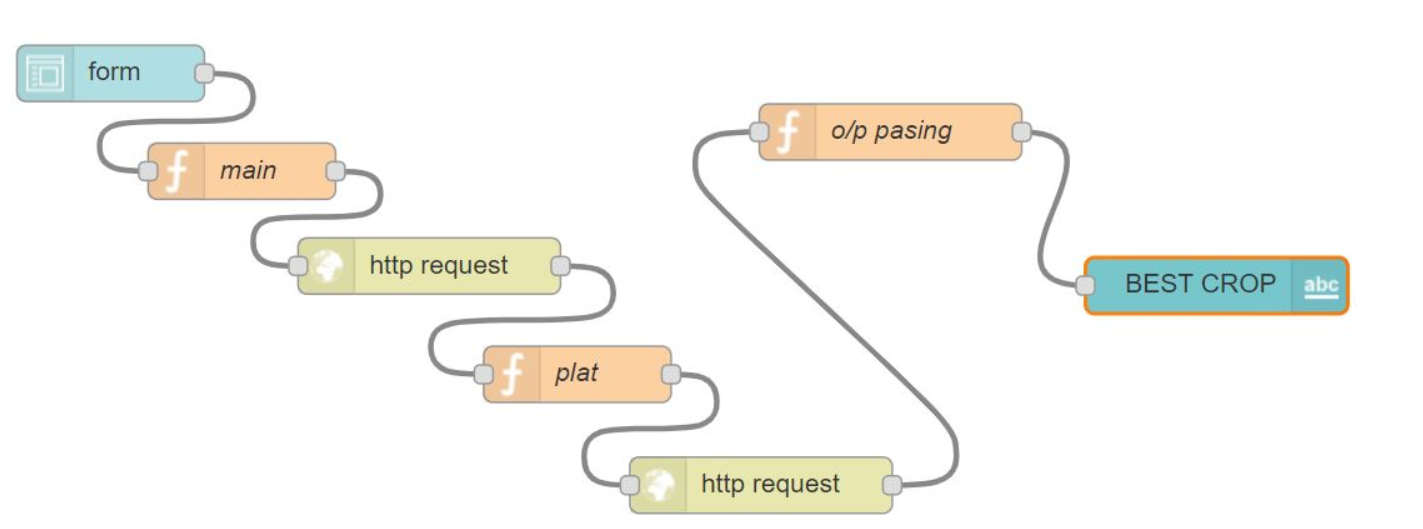




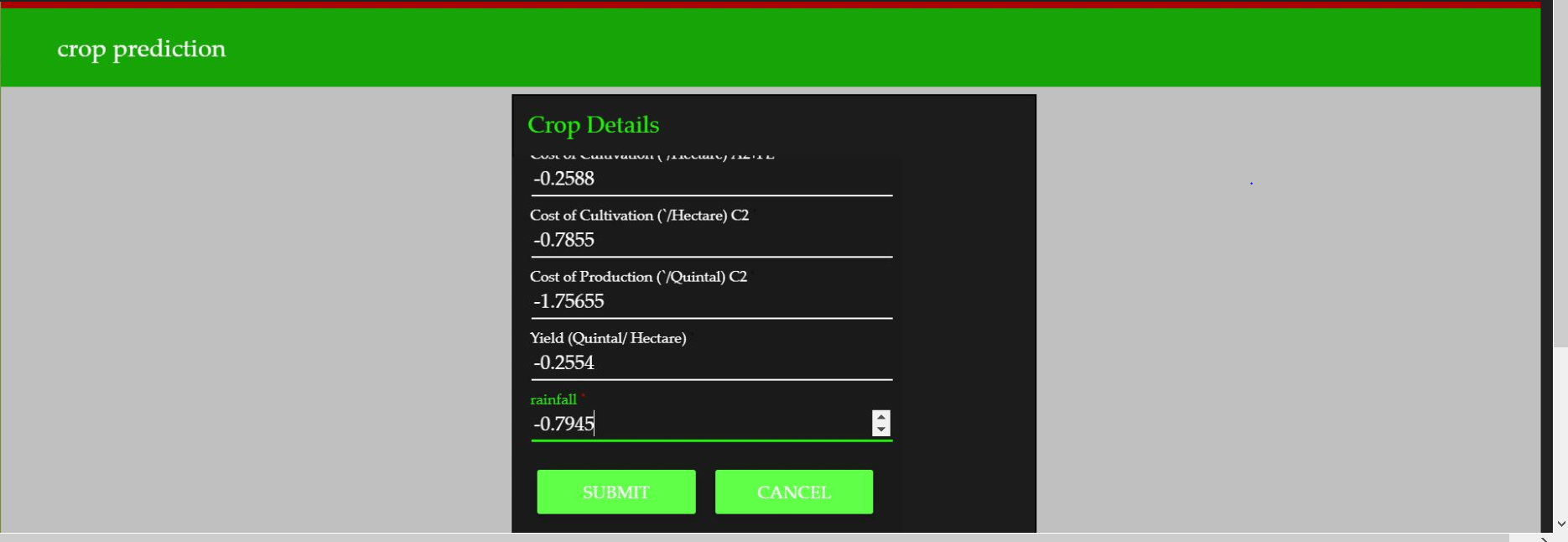
ROC CURVE:



Data Clustering In the clustering step, the preprocessed data was clustered using k-means clustering algorithm. Kmeans is the popular clustering algorithm in this category. The k value refers to the number of clusters in which the data are partitioned. Clusters are represented by their centers. The basic idea is that each sample should be closer to the center of its own cluster. If this is not verified, then the partition is modified, until each sample is closer to the center of the cluster it belongs to. The distance function between samples plays an important role, since a sample can migrate from a cluster to another one based on the values provided by the distance function.



5.PREDICTION:



6. CONCLUSION:

Crop yield prediction is still remaining as a challenging issue for farmers. The aim of this research is to propose and implement a rule based system to predict the crop yield production from the collection of past data. This has been achieved by applying association rule for crop yielding prediction agriculture data from 2000 to 2012