**CHAPTER 1**

# INTRODUCTION

Data Analysis is a process of cleaning inspection, data modeling with the objective of finding useful information and conclusions. In order to extract some pattern, it is a process of analyzing, extracting and predicting meaningful information from huge data. Farmers use this method to collect their customer's raw data for useful information. This analysis can also be used in the field of Agriculture. Most farmers were depend on on their long-terms experiences in the field on particular crops to expect a higher yield in the next harvesting period But still they don’t get worth price of the crops. It is mostly happens due to improper irrigation or inappropriate crops selection or also sometimes the crop yield is less than that of expected. Agricultural researchers insist on the need for an efficient mechanism to predict and improve the crop growth and Majority of research works in agriculture focus on biological mechanisms to identify crop growth and improve its yield. The outcome of crop yield primarily depends on parameters such as variety of crop, seed type and environmental parameters such as sunlight (Temperature), soil (ph), water (ph), rainfall and humidity. By analysing the soil and atmosphere at particular region best crop in order to have more crop yield and the net crop yield can be predict. This prediction will help the farmers. To choose appropriate crops for their farm according to the soil type, temperature, humidity, water level, spacing depth, soil PH, season, fertilizer and months. Crop yield estimation is a difficult task since it is affected by various factors such as genetic potential of crop cultivar, soil, weather, cultivation practices (date of sowing, amount of irrigation and fertilizer, etc.) and biotic stress . Several methods of crop yield estimation have been developed such as statistical, agrometeorological, empirical, biophysical, and mechanistic.

India is a highly populated country and randomly change in the climatic conditions need to secure the world food resources. Framers face serious problems in drought conditions. Type of soil plays a major role in the crop yield. Suggesting the use of fertilizers may help the farmers to make the best decision for their cropping situation . The number of studies Information and Communication Technology (ICT) can be applied for prediction of crop yield .By the use of Data Mining, we can also predict the crop yield. By fully analyse the previous data we can suggest the farmer for a better crop for the better yield.

Smart agriculture is the way of conveying information from traditional farmers to the educated farmers. To obtain estimates of aggregate physical production functions for the yields of various crops in specified states, considering various technological factors and a newly developed weather index as inputs. Regression and coefficient of determination analysis along with Average Error rate were carried out to make a decent comparison between our actual result which is called target and prediction model that is friendly interface for farmers, which gives the analysis of rice production based on available data. Different Data mining techniques were used to predict the crop yield for maximizing the crop productivity Accurate and timely monitoring of agricultural crop conditions and estimating potential crop yields are essential processes for operational programs Because of the importance of predicting crop yield, the purpose of this study is to apply several forecasting methods for evaluating crop yield estimates in Ghana. Crop yield forecasting, which provides information for decision Makers.

## 1.1 Existing system

INDIA is a highly populated country and randomly change in the climatic conditions need to secure the world food resources. Framers face serious problems in drought conditions. Type of soil plays a major role in the crop yield. Suggesting the use of fertilizers may help the farmers to make the best decision for their cropping situation [1]. The number of studies Information and Communication Technology (ICT) can be applied for prediction of crop yield [2]. By the use of Data Mining, we can also predict the crop yield. By fully analyse the previous data we can suggest the farmer for a better crop for the better yield [3].

## 1.2Proposed System

This project is used to predict the crop yield and suitable crop by considering the information such as soil type, temperature, humidity, season, fertilizer and months. The system provides easier and faster access to All the basic information regarding the District, rainfall, area under irrigation, crop, season yield, fertilizers used through which user can analysis the crop and also select the option of Prediction where he can select the crop production parameters to get the suitable crop for his farm. This system provides simple visualization so that user can understand and analysis things in easy way.

## 1.3 Motivation

Farming is the main occupation of India. About 70 percent of primary and secondary business is based on farming. So for the betterment of farming many farmers have started using the new technologies and methods. But people don’t have awareness about the cultivation of the crops in a right time and at a right place. In this case an idea to identify the suitability of crops and yield based on various factors that affect the production can increase the quality and the yield of crops, thereby increase the economic growth and attain profitability.

Agriculture is a business with risk and reliable crop yield prediction is vital for decisions related to agriculture risk management. The vision of meeting world’s food demands for the increasing population throughout the world is becoming more important in these recent years. Eventually, helps in achieving ZERO hunger. Predictions could be used by crop managers to minimize losses when unfavourable conditions may occur.

## 1.4 Objectives of the work

The main objective of this project is to predict crop yield by using ensemble learning techniques with higher accuracy.

## 1.5 Key features

There is a main machine learning algorithm that is used to determine whether farmers get their yield or not. In the present system SVM algorithm is used as classifier which proves to give the highest accuracy among the other algorithms. Instances can be classified by more than one output. There is a broad survey conducted on the type of machine learning algorithm used, based on which the yield of crop is predicted. At a more fundamental level, it is evident that the machine learning is also helping to improve our basic understanding of crop prediction.

## 1.6 Organization of the project report

This report is divided into 5 parts:

* In the chapter 1 we have discuss about the introduction about domain, introduction about this topic, scope of the project.
* In the chapter 2 we have discussed about the literature review and its merits and demerits related to this topic.
* In the chapter 3 we have discuss about the system requirements and specifications.
* In the chapter 4 we have discuss about the system architecture and design.
* In the chapter 5 we have discuss about the system implementation and modules with descriptions.

**CHAPTER 2**

# LITERTURE SURVEY

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, then next step is to determine which operating system and language can be used for developing the tool.

## 2.1 Rice yield estimation at pixel scale using relative vegetation indices from unmanned aerial systems (2019)

**Author**: Feilong Wang, Fumin Wang\*,YaoZhang,Jinghui Hu

**Publications**: “Open Fund of Laboratory of Target background conditions in field scale, this

Microwave Properties” (2018KF02) “National Natural Science Foundation of China”

(41871328) and “National Key Research and Development Plan of china”

**Work carried out in the paper**:

Timely and accurate prediction of rice yield information is closely related to the people's livelihood, Satellite remote sensing provides the possibility for large-scale crop yield estimation, but they are usually limited by spatial and spectral resolution. Unmanned Aerial Vehicles (UAV) remote sensing with hyper spectral sensors can obtain high spatial temporal resolution and hyper spectral images on demand. Generally, time-series Vegetation Indices (VIs) are used for estimating grain yield. So the differences between vegetation indices may include the effects induced from external condition, which will pose a negative effect on the accuracy of crop yield estimation. Therefore, in this study, the relative vegetation index and relative yield were proposed and used to estimate rice yield at pixel scale. And the optimal growth stages for crop yield estimation would also be determined. Then the best RNDVI at different growth stages were determined for rice yield estimation. Finally, different combinations of growth stages were tested to obtain the optimal combinations for yield estimation. These models were validated at pixel scale using the measured yields. Results shows that the yield estimations at pixel scale with relative vegetation indices were acceptable. In the study, a yield estimation method with relative vegetation indices is proposed and the optimal growth stage combinations for rice yield estimation were determined. This study explores the possibility of yield estimation at pixel scale using hyper spectral images from UAV platform, which will further improve the method system for remote sensing of yield estimation.

**Merits:** In the study, a yield estimation method with relative vegetation indices is proposed and the optimal growth stage combinations for rice yield estimation were determined.

**Demerits:** Applicable only for single data.

**2.2 Use Of Deep Neural Networks For Crop Yield Prediction: A Case Study Of Soybean Yield in Lauderdale County, Alabama, USA (2019)**

**Author:** AnılSuatTerliksiz, D. TurgayAltılar

## Publications: IEEE

**Work carried out in the paper**:

World population is constantly increasing and it is necessary to have sufficient crop production. Deep learning gains importance on crop monitoring, crop type classification and crop yield estimation applications with the recent advances in image classification using deep Convolutional Neural Networks (CNN). Traditional crop yield prediction approaches based on remote sensing consist of classical Machine Learning methods such as Support Vector Machines and Decision Trees. CNN and Long-Short Term Memory Network (LSTM) are deep neural network models that are proposed for crop yield prediction recently. This study focused on soybean yield prediction of Lauderdale County, Alabama, USA using 3D CNN model that leverages the spatiotemporal features. The yield is provided from USDA NASS

Quick Stat tool for years 2003-2016. The satellite data used is collected from NASA’s MODIS land products surface reflectance, land surface temperature and land surface temperature via Google Earth Engine. The root mean squared error (RMSE) is used as the evaluation metric in order to be able to compare the results with other methods that generally use RMSE as the evaluation metric.

**Merits:** This project mainly delays with collection of the land product surface reflectance land surface temperature via Google earth engine.

**Demerits:** This project fails in predicting type of crop.

## 2.3 Prediction of major crop yields of Tamil nadu using K-means and Modified KNN (2018)

**Author:** Mr A Suresh, Dr. P. Ganesh Kumar, Dr.M.Ramalatha

**Publications:** Proceedings of the International Conference on Communication and

Electronics Systems (ICCES) IEEE Xplore

**Work carried out in the paper**:

Agriculture is the principal source of livelihood for more than 40 percent of the population of this state. According to Food and Agricultural Organization (FAO) researchers, between 2010 and 2050 the world population will increase by one third. The demand for crop production will increase by 60percent higher than the current production. Hence prediction plays a major role to find out the demand of crop production for maximizing the yield. For that in this paper we propose a prediction method for the major crops of Tamil nadu using Kmeans and Modified K Nearest Neighbor (KNN). MatlabandWEKA are used as the tool for clustering and classification respectively. The number result shows that our method is better than traditional data mining approach.

**Merits:** It finds the demand of crop production for maximizing yield

**Demerits:** Less efficiency

## 2.4 Rice crop yield prediction in India using support vector machines (2016)

**Author:** NiketaGandhi; Leisa J. Armstrong; OwaizPetkar; Amiya Kumar Tripathy

**Publications:** 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)

**Work carried out in the paper**:

Food production in India is largely dependent on cereal crops including rice, wheat and various pulses. The sustainability and productivity of rice growing areas is dependent on suitable climatic conditions. Variability in seasonal climate conditions can have detrimental effect, with incidents of drought reducing production. Developing better techniques to predict crop productivity in different climatic conditions can assist farmer and other stakeholders in better decision making in terms of agronomy and crop choice. Machine learning techniques can be used to improve prediction of crop yield under different climatic scenarios. This paper presents the review on use of such machine learning technique for Indian rice cropping areas. This paper discusses the experimental results obtained by applying SMO classifier using the WEKA tool on the dataset of 27 districts of Maharashtra state, India. The dataset considered for the rice crop yield prediction was sourced from publicly available Indian Government records. The parameters considered for the study were precipitation, minimum temperature, average temperature, maximum temperature and reference crop evapotranspiration, area, production and yield for the Kharif season (June to November) for the years 1998 to 2002. For the present study the mean absolute error (MAE), root mean squared error (RMSE), relative absolute error (RAE) and root relative squared error (RRSE) were calculated. The experimental results showed that the performance of other techniques on the same dataset was much better compared to SMO.

**Merits:** It provides a better result by using SVM when compared to SMO classifier.

**Demerits:** Large datasets cannot be predicted.

## 2.5 Regional Winter Wheat Maturity Date Prediction Using Remote Sensing-Crop Model Data Assimilation and Numerical Weather Prediction (2018)

**Author:** XinranGao; Jianxi Huang; Hongyuan Ma; Wen Zhou; Dehai Zhu

**Publications**: 7th International Conference on Agro-geoinformatics (Agro-geoinformatics) **Work carried out in the paper**:

Optimizing harvesting schedules requires a method for maturity date prediction, to avoid the influence of adverse weather and prevent the decline of crop yield or quality due to inappropriate harvest schedule. However, most prediction models are statistical-based thus are not suitable for regional application, and remote sensing-based models lacked predictability. We presented a framework that assimilated leaf area index (LAI) derived from Moderate Resolution Imaging Spectro radiometer (MODIS) into World Food Studies

(WOFOST) crop growth model, and forecast meteorological data from THORPEX Interactive Grand Global Ensemble (TIGGE) was used as weather data input for the future periods. We selected the winter wheat planting area in Henan Province as study area and recalibrated WOFOST model based on observation data from agro-meteorological sites. A cost function based on normalization was constructed to quantify the difference between simulated LAI and MODIS LAI products. First the MODIS LAI profile was smoothed by Savitzky-Golay (S-G) filter, and then these two LAI profiles were normalized to keep their trend information. Then we selected parameters in WOFOST model that are sensitive to maturity date as optimization parameters, such as emergence date (IDEM), effective temperature sum from emergence to anthesis (TSUM1) and effective temperature sum from anthesis to maturity (TSUM2). These parameters have significant differences between years and no obvious spatial and temporal patterns. By means of Shuffled Complex Evolution method developed at the University of Arizona (SCE-UA) algorithm, we simulated in each pixel in the study area and retrieved the optimal parameters set of this pixel. Then we run WOFOST by this optimal parameter set to simulate the growth and development of winter wheat. Moreover, we transformed TIGGE data into the CABO-format weather file to drive WOFOST simulating winter wheat growth in the next 16 d and obtained a spatial distribution of winter wheat. Moreover, we transformed TIGGE data into the CABO-format weather file to drive WOFOST simulating winter wheat growth in the next 16 d and obtained a spatial distribution of winter wheat maturity date in the study area. Comparing the forecasting date with the observed date from agro-meteorological sites, it demonstrated that this method had substantial accuracy in predicting regional maturity date with correlation coefficient (R2) of 0.90 and the root mean square error (RMSE) was 1.93 d. Besides that, the distribution map of maturity prediction showed obvious spatial variability. This method can remedy the shortages of poor predictability and lacking regional differences in most previous methods, and it provides a reference for the future study of crop maturity prediction at a regional scale with longer forecast period.

**Merits:** This method can remedy the shortages of poor predictability and lacking regional differences in most previous methods, and it provides a reference for the future study of crop maturity prediction at a regional scale with longer forecast period.

**Demerits:** Result generated are in accurate.

## 2.6 Multiple Crop Yield prediction using dual-polarimetricTerraSAR-X stripmap imagery (2009)

**Author:** TishampatiDha; Doug Gray; Carl Menges

**Publication:** IEEE International Geoscience and Remote Sensing Symposium

**Work carried out in the paper**:

This paper presents the results of an experiment carried out to relate the yield from various crops to TerraSAR-X dual polarimetric imagery. X-band wavelength has higher sensitivity to smaller crop structures, especially stem and head density making it suitable for relating yield to backscatter. The coherent dual-polarimetric mode of TerraSAR-X was also used to emphasize the volume scattering through dual-polarimetric entropy/alpha decomposition. Good correlations to yield data as gathered by harvester telemetry were obtained.

**Merits:** X-band wavelength has higher sensitivity to smaller crop structures, especially stem and head density making it suitable for relating yield to backscatter.

**Demerits:** Failed in capturing multiple image of crop

## 2.7 Remote sensing indicators for crop growth monitoring at different scales (2011)

**Author**: ZongnanLi; Zhongxin Chen

**Publications**: IEEE International Geoscience and Remote Sensing Symposium.

**Work carried out in the paper**:

Crop growth monitoring is critical in yield estimation and prediction. In this paper, the authors investigated several indicators for crop growth monitoring by remote sensing at different scales. The experiments were conducted in a study area in Hebei province in North China Plain. The target crop in this research is winter wheat, which is one of the important grain crops in China. The study at canopy scale is based on field experiment with different fertilizer supply treatments to winter wheat to get different crop growth gradients. The canopy spectra and LAI data of different crop growth situation were collected and analysed. The correlation coefficients between vegetation indices and LAI were calculated. The best growth monitoring indicators at canopy scale for each phonological stage were determined by the correlation coefficients. The best indicator in early elongation stage is SAVI (L=0.3). While in jointing stage, the best indicator is SAVI (L=0.2). In the heading and milk stages, the best indicator is NDVI. At the regional scales, the multi-spectral remote sensing data from HJ-IA satellite were used to study the crop growth monitoring indicators. The performances of the vegetation indices at different spatial scales for crop growth monitoring were studied in different phonological stages by comparing the correlation coefficients of vegetation indices and LAI data. The result shows that the vegetation indices which have a function to decrease soil-effect are suitable for the region in different crop coverage in early elongation stage and for the region in low crop coverage, but they have limitations in dense crop coverage while in booting and heading stages.

**Merits:** The result shows that the vegetation indices which have a function to decrease soileffect are suitable for the region in different crop coverage in early elongation stage and for the region in low crop coverage

**Demerits:** This project focuses only on one particular crop rather than many other crops.

## 2.8 Assessing relevant climate data for agricultural applications (2012)

**Author:** J. Ramirez-Villegas and A. Challinor **Publications:** Agricultural Forest Meteorology

**Work carried out in the paper**:

Climate change is expected to substantially reduce agricultural yields, as reported in the by the Intergovernmental Panel on Climate Change (IPCC). In Sub-Saharan Africa and (to a lesser extent) in South Asia, limited data availability and institutional networking constrain agricultural research and development. Here we performed a review of relevant aspects in relation to coupling agriculture–climate predictions, and a three-step analysis of the importance of climate data for agricultural impact assessment. First, using meta-data from the scientific literature we examined trends in the use of climate and weather data in agricultural research, and we found that despite agricultural researchers’ preference for field-scale weather data (50.4% of cases in the assembled literature), large-scale datasets coupled with weather generators can be useful in the agricultural context. Using well-known interpolation techniques, we then assessed the sensitivities of the weather station network to the lack of data and found high sensitivities to data loss only over mountainous areas in Nepal and Ethiopia (random removal of data impacted precipitation estimates by ±1300 mm/year and temperature estimates by ±3 ◦C). Finally, we numerically compared IPCC Fourth Assessment Report (4AR) climate models’ representation of mean climates and inter annual variability with different observational datasets. Climate models were found inadequate for field-scale agricultural studies in West Africa and South Asia, as their ability to represent mean climates and climate variability was limited: more than 50% of the country-model combinations showed.

**Merits:** Using well-known interpolation techniques, we then assessed the sensitivities of the weather station network to the lack of data and found high sensitivities to data loss only over mountainous areas.

**Demerits:** Less effective when compared to other.

## 2.9 Crop Syst, a Cropping Systems Simulation Model: Water/Nitrogen Budgets and Crop Yield (1994)

**Author:** Claudio O. Stockle, Steve A. Martin &Gaylon S. Campbell

**Publications:** Agricultural Systems

**Work carried out in the paper**:

In agriculture, water and nitrogen are two critical resources for growing a crop. However, their management cannot be analysed independently of weather, soil characteristics, field hydrology, crop characteristics, crop rotation, and management factors. This paper describes the water, nitrogen, and crop growth components of Crop Syst, a comprehensive cropping systems simulation model, and provides preliminary verification of these components. The water budget of the model properly describes crop water use. Predicted nitrogen contents throughout the soil profile did not exactly match the measured values from leaching experiments, but they did follow the general trends of the data. The agreement between simulated and observed biomass and yield of corn, winter wheat and spring wheat grown in two locations with a total of 77 data points was good as shown by several statistical indicators. Based on this preliminary validation, Crop Syst appears promising as a tool to analyse management practices for water and nitrogen. Additional validation of model components, including a wide range of crops and conditions, should be conducted in the future.

**Merits:** This project predicts the budget for water and nitrogen use.

**Demerits:** Predicted nitrogen content contents throughout the soil profile did not exactly match the measured values from leaching experiments.

**CHAPTER 3**

# SYSTEM REQUIREMENT SPECIFICATION

A software requirements specification (SRS) – a[requirements specification](http://en.wikipedia.org/wiki/Requirements_specification)for a[software system](http://en.wikipedia.org/wiki/Software_system)– is a complete description of the behaviour of a system to be developed. In addition to a description of the[software functions,](http://en.wikipedia.org/wiki/Functional_requirements) the SRS also contains[non-functional requirements.](http://en.wikipedia.org/wiki/Non-functional_requirements) Software requirements are a sub-field of[software engineering](http://en.wikipedia.org/wiki/Software_engineering)that deals with the elicitation, analysis, specification, and validation of[requirements](http://en.wikipedia.org/wiki/Requirements)for[software.](http://en.wikipedia.org/wiki/Software)

## 3.1 FEASIBILITY STUDY

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

Three key considerations involved in the feasibility analysis are

* Economical Feasibility
* Technical Feasibility
* Social Feasibility

### Economical Feasibility

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

### Technical Feasibility

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

### Social Feasibility

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

## 3.2 FUNCTIONAL REQUIREMENT

The particular necessities are user interface. The outside clients are farmers. Every one of the farmers can utilize this product for ordering and looking.

* Hardware Interface: The outside equipment interface utilized for ordering and looking is PCs of farmers. The PCs might be portable PCs with remote LAN as the web association gave will be remote.
* Software Interfaces: The working Frameworks can be any rendition of windows.
* Performance Prerequisites: The PCs utilized must be Pentium 4 machine with the goal that they can give ideal execution of the item.

## 3.3 NON-FUNCTIONAL REQUIREMENT

Non utilitarian necessities are the capacities offered by the framework. It incorporates time imperative and requirement on the advancement procedure and models. The non-useful prerequisites are as per the following:

* Speed: The framework ought to prepare the given contribution to yield inside fitting time.
* Ease of utilization: The product thought to be easy to understand. At that point the clients can utilize effortlessly, so it doesn’t require much preparing time.
* Reliability: The rate of disappointments ought to be less then just the framework is more solid.
* Portability: It thought to be anything but difficult to actualize in any framework.

## 3.4 TOOLS AND TECHNOLOGY DETAILS

Hardware requirements are:

* i3 Processor Based Computer or higher
* RAM Memory: 4/8 GB
* Hard Drive: 50 GB
* Internet Connection

**SOFTWARE REQUIREMENTS:**

* Operating system : Windows 7 or higher
* Coding Language : Python 3.7
* Pycharm

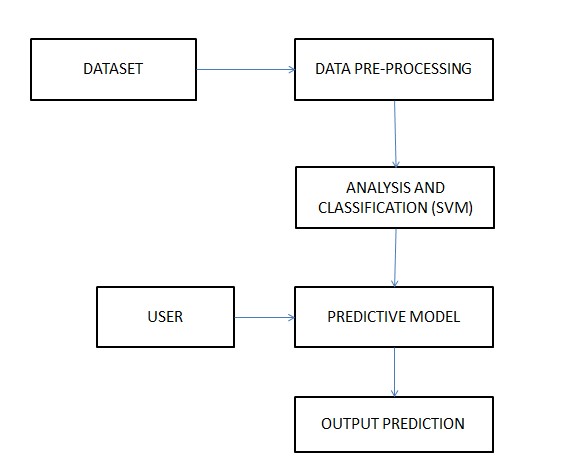
**CHAPTER 4**

# SYSTEM DESIGN

System design is the process of the defining the architecture, components, modules, interfaces, and data for a system to satisfy specified requirements .Systems design could be seen as the application of systems theory to product development. Object- oriented analysis and methods are becoming the most widely used methods for computer systems design. Systems design is therefore the process of defining and developing systems to satisfy specified requirements of the user. The UML has become the standard language in object oriented analysis and design

## 4.1 System Architecture Diagram

System architecture is a conceptual model that defines the structure and behavior of the system. It comprises of the system components and the relationship describing how they work together to implement the overall system.



**Fig 4.1: System design for proposed system**

## 4.2 Input/output Oriented Design

Input: 10year crop data

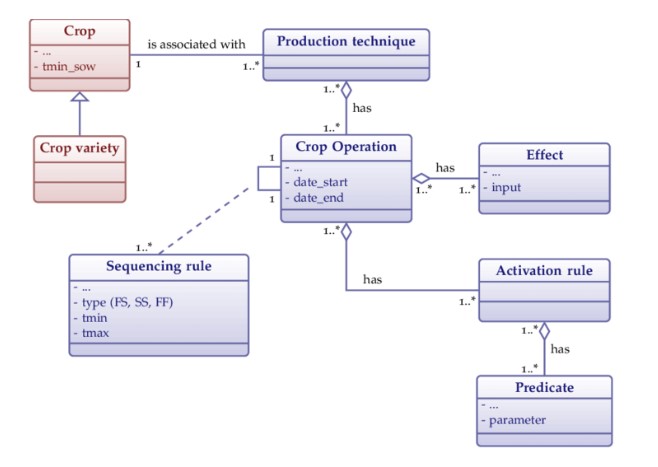
Output: Crop yield prediction of this year.

## 4.3 Object Oriented Design

During the detailed phase, the view of the application developed during the high level design is broken down into modules and programs. Logic design is done for every program and then documented as program specifications. For every program, a unit test plan is created.

### 4.3.1 Class diagram for proposed System

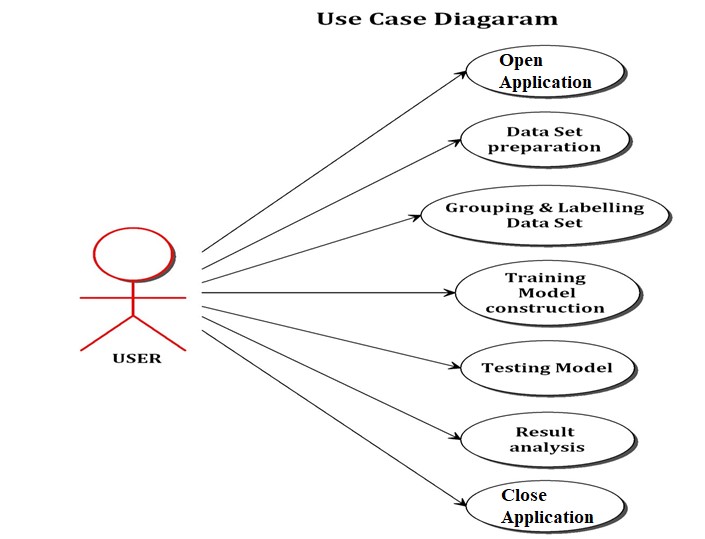
Class diagram is a concept of production technique and related classes. It contains attributes in its classes. For simplification, we did not include all attributes describing each classes.



**Fig 4.2: Class diagram**

### 4.3.2 Use case diagram for proposed System

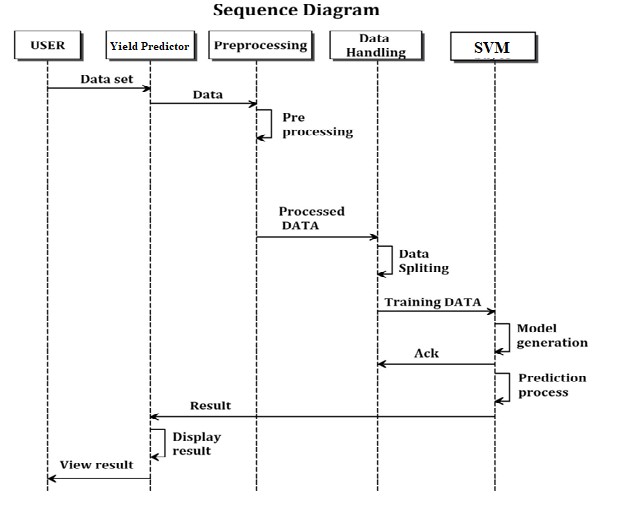
A use case diagram at its simplest is a representation of a user's interaction with the system that shows the relationship between the user and the different use cases in which the user is involved. A use case diagram can identify the different types of users of a system and the different use cases and will often be accompanied by other types of diagrams as well. While a use case itself might drill into a lot of detail about every possibility, a use case diagram can help provide a higher-level view of the system. It has been said before that "Use case diagrams are the blueprints for your system". They provide the simplified and graphical representation of what the system must actually do.



**Fig 4.3: Use case diagram**

### 4.3.3 Sequence diagram for proposed System

The user provides the dataset to predict the yield, the data is pre-processed several times in pre-processing, the pre-processed data is sent to data handling, in data handling the data splits, after splitting the training data is sent to SVM, in SVM the data is generates model, SVM acknowledges the data handling, then SVM predicts the process, the result is sent to yield predictor, yield predictor displays the result and finally the user will view the result.

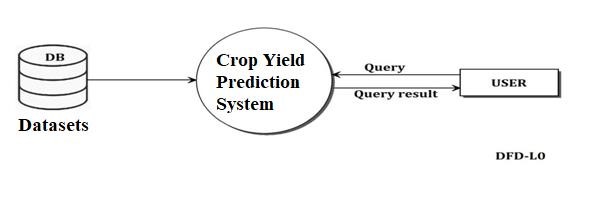


**Fig 4.4: Sequence diagram**

### 4.3.4 Dataflow diagram for proposed System

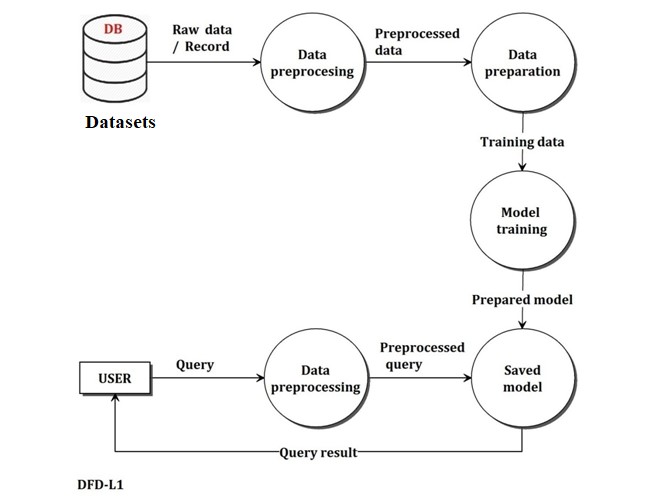
A dataflow diagram is a graphical representation of the "flow" of data through an information system, modelling its process aspects. A DFD is often used as a preliminary step to create an overview of the system without going into great detail, which can later be elaborated. DFDs can also be used for the visualization of data processing. A DFD shows what kind of information will be input to and output from the system, how the data will advance through the system, and where the data will be stored.

Level 0:



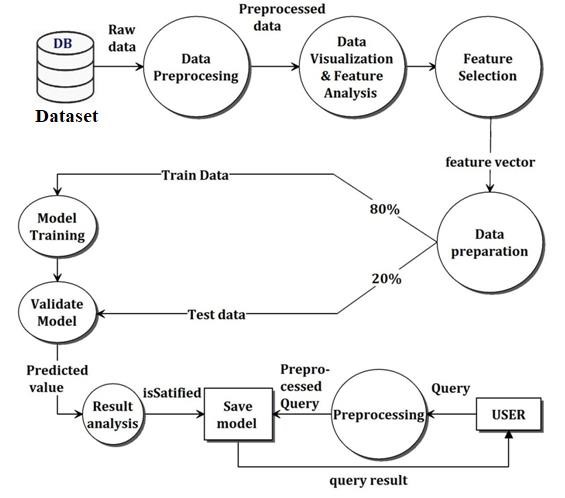
#### Fig 4.5 Dataflow diagram Level 0

Level 1:



#### Fig 4.6: Dataflow diagram Level1

Level 2:

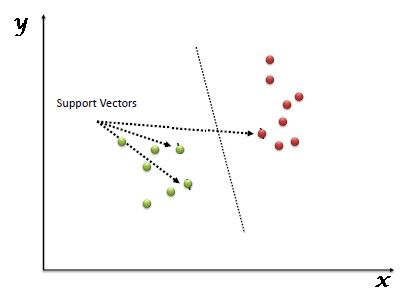


**Fig 4.7: Dataflow diagram Level 2**

## 4.4 Algorithm

“Support Vector Machine” (SVM) is a supervised[machine learning algorithm](https://courses.analyticsvidhya.com/courses/introduction-to-data-science-2?utm_source=blog&utm_medium=understandingsupportvectormachinearticle)which can be used for both classification and regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in ndimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well (look at the below snapshot).

Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyper-plane/ line).



**Fig 4.8: SVM**

Above, we got accustomed to the process of segregating the two classes with a hyper-plane.

**CHAPTER 5**

# SYSTEM IMPLEMENTATION

This project is used to prediction of crop yield and suitable crop by considering the information such as soil type, temperature, humidity, season, fertilizer and months. The system provides easier and faster access to All the basic information regarding the District, rainfall, area under irrigation, crop, season yield, fertilizers used through which user can analysis the crop and also select the option of Prediction where he can select the crop production parameters to get the suitable crop for his farm. This system provides simple visualization so that user can understand and analysis things in easy way.

## 5.1 Modules

* Data preprocessing
* Dataset classification
* Predictive model

### 5.1.1 Data pre-processing

The crop dataset of particular district is loaded to the system. It contains a set of raw data. Data is gathered from different sources, it is connected in row format. Data processing is a technique that is used to convert the raw data into a clean dataset. It refers to transformation applied to our data before feeding it to the algorithm. Hear the raw data in the crop data is cleaned and the metadata is appending to it by removing the things which are converted to the integer. So, the data is easy to train. Hear all the data. In this pre-processing, we first load the metadata into this and then this metadata will be attached to the data and replace the converted data with metadata. Then this data will be moved further and remove the unwanted data in the list and it will divide the data into the train and the test data For this splitting of the data into train and test we need to import train\_test\_split which in the scikit-learn this will help the pre-processed data to split the data into train and test according to the given weight given in the code. The division of the test and train is done in 0.2 and 0.8 that is 20 and 80 percent respectively.

### 5.1.2 Estimating algorithm

We use a simple vector machine algorithm. SVM can be used for both classification and regression challenges. This is a training phase. Here training data and labeled data are fed as input to train the classifier SVM learns existing data and classifies the unlabeled data. Different Data mining techniques were used to predict the crop yield for maximizing the crop productivity Accurate and timely monitoring of agricultural crop conditions and estimating potential crop yields are essential processes for operational programs Because of the importance of predicting crop yield, the purpose of this study is to apply several forecasting methods for evaluating crop yield.

### 5.1.3 Predictive model

Predictive model is a testing phase. Farmer give an input as a new data. New data is compared with the existing trained dataset and predict the proper output. Here we predict the crop yield and efficient use of fertilizer.

**CHAPTER 6**

**SYSTEM TESTING**

**6.1 Testing**

This chapter gives the outline of all testing methods that are carried out to get a bug free system. Quality can be achieved by testing the product using different techniques at different phases of the project development. The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components sub-assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

**6.2 Manual testing**

Manual testing is a testing process that is carried out manually in order to find defects without the usage of tools or automation scripting. A test plan document is prepared that acts as a guide to the testing process in order to have the complete test coverage.Testing is an integral part of software development. Testing process certifies whether the product that is developed compiles with the standards that it was designed to. Testing process involves building of test cases against which the product has to be tested.

**6.3 Unit testing**

Unit testing is performed for testing modules against detailed design. Inputs to the processare usually compiled modules from the coding process. Each modulesare assembled into a larger unit during the unit testingprocess.

Testing has been performed on each phase of project design and coding. We carry out the testing of module interface to ensure theproper flow of information into and out of the program unit while testing. We make sure that the temporarily stored data maintains its integrity throughout the algorithm's executionby examining the loc al data structure. Finally, all error-handling paths are alsotested.

**6.4 Integration testing**

`In my system the following testing is done

-When a record is inserted in the front end,it gets saved in the back end .

- Modification made are reflected in the backend .

-Deletion of a record will delete individual tables are tested, hence unit testing done and All forms are linked and checked therefore integreation testing is done.

**6.5 System testing**

We usually perform system testing to find errors resulting from unanticipatedinteraction between the sub-systemandsystemcomponents.Softwaremustbe tested to detect and rectify all possible errors oncethesourcecode isgenerated before delivering it to the customers. For finding errors, series of test cases must be developed which ultimately uncover all the possibly existingerrors. Different software techniques can be used for thisprocess. Thesetechniques providesystematic guidance for designing testthat

* + - Exercise the internal logic of the softwarecomponents,
    - Exercise the input and output domains of a program to uncover errors in program function, behavior andperformance.

We test the software using two methods:

White Box testing: Internal program logic is exercised using this test case design techniques.

Black Box testing: Software requirements are exercised using this test case design techniques.

Finally system testing is done on deployment.

Unit testing:

|  |  |
| --- | --- |
| Sl # Test Case : - | UTC-1 |
| Name of Test: - | Uploading dataset. |
| Items being tested: - | Tested for uploading dataset. |
| Sample Input: - | Upload trained data |
| Expected output: - | dataset should upload properly |
| Actual output: - | upload successful |
| Remarks: - | Pass. |

|  |  |
| --- | --- |
| Sl # Test Case : - | UTC-2 |
| Name of Test: - | Detecting soil dataset |
| Items being tested: - | Test for all the dataset |
| Sample Input: - | Tested for all the dataset which you have trained.. |
| Expected output: - | Classified soil result should be displayed |
| Actual output: - | Should Display soil classification result. |
| Remarks: - | result displayed |

Integration Testing:

|  |  |
| --- | --- |
| Sl # Test Case : - | ITC-1 |
| Name of Test: - | Working of Choose File option |
| Item being tested: - | User convenience to access dataset stored |
| Sample Input: - | Click and select the dataset |
| Expected output: - | Should open dataset which we have trained. |
| Actual output: - | Selected dataset should load |
| Remarks: - | Pass. |

|  |  |
| --- | --- |
| Sl # Test Case : - | ITC-2 |
| Name of Test: - | Working of soil classification and Displaying crop yield |
| Item being tested: - | Selecting soil classification dataset and verifying result. |
| Sample Input: - | Click and select dataset |
| Expected output: - | Should show exact output. |
| Actual output: - | Soil classification output should be displayed |
| Remarks: - | Pass. |

System testing:

|  |  |
| --- | --- |
| Sl # Test Case : - | STC-1 |
| Name of Test: - | System testing in various versions of OS |
| Item being tested: - | OS compatibility. |
| Sample Input: - | Execute the program in windows XP/ Windows-7/8 |
| Expected output: - | Performance is better in windows-7/10 |
| Actual output: - | Same as expected output, performance is better in windows-7/10 |
| Remarks: - | Pass |

**CHAPTER 7**

**RESULTS**

This section describes the output of the “SOIL CLASSIFICATION”. The snapshots are shown below for each module.

**SNAPSHOTS**

**(Place the snapshots of your project)**

**CONCLUSION**

Agriculture is the backbone of counties like India. However, the usage of technology towards agriculture is to be given paramount importance towards preclSlonagriculture.This paper proposes a system which will help farmers to have an idea of yield estimates based on weather parameters and area under cultivation Using this farmer can make decisions on whether to grow that particular crop or go for alternate crop in case yield predictions are unfavorable.This research work can be enhancing to the next level. We can build a recommender system of agriculture production and distribution for farmer. By which farmers can make decision in which season which crop should sow so that they can get more benefit. This system is work for structured dataset. In future we can implement data independent system also. It means format of data whatever, our system should work with same efficiency.

**REFERENCES**

[1] Ponce-Guevara, K. L., Palacios-Echeverria, J. A., Maya-Olalla, E.,DominguezLimaico, H. M., Suarez-Zambrano, L. E., Rosero-Montalvo, P.D., Alvarado-Perez, J. C. (2017). GreenFarm-DM: A tool for analyzing vegetable crops data from a greenhouse using data mining techniques (First trial). 2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM).

[2] Jheng, T.-Z., Li, T.-H., Lee, C.-P. (2018). Using hybrid support vector regression to predict agricultural output. 2018 27th Wireless and Optical Communication Conference (WOCC).

[3] Manjunatha, M., Parkavi, A. (2018). Estimation of Arecanut Yield in Various Climatic Zones of Karnataka using Data Mining Technique: A Survey. 2018 International Conference on Current Trends Towards Con- verging Technologies (ICCTCT).

[4] Shakoor, M. T., Rahman, K., Rayta, S. N., Chakrabarty, A. (2017). Agricultural production output prediction using Supervised Machine Learning techniques.2017 1st International Conference on Next Generation Computing Applications (NextComp).

[5] Grajales, D. F. P., Mejia, F., Mosquera, G. J. A., Piedrahita, L. C.,Basurto,C. (2015). Crop-planning, making smarter agriculture with climate data. 2015 Fourth International Conference on Agro-Geoinformatics (Agro-Geoinformatics).

[6] Shah, P., Hiremath, D., Chaudhary, S. (2017). Towards development of spark based agricultural information system including geo-spatial data. 2017 IEEE International Conference on Big Data (Big Data).

[7] Afrin, S., Khan, A. T., Mahia, M., Ahsan, R., Mishal, M. R., Ahmed, W.,Rahman, R. M. (2018). Analysis of Soil Properties and Climatic Data to Predict Crop Yields and Cluster Di\_erent Agricultural Regions of Bangladesh.2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS).

[8] Sekhar, C. C., Sekhar, C. (2017). Productivity improvement in agriculture sector using big data tools. 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC).

[9] Sahu, S., Chawla, M., Khare, N. (2017). An e\_cient analysis of crop yield prediction using Hadoop framework based on random forest approach. 2017 International Conference on Computing, Communication and Automation (ICCCA).

[10] Garg, A., Garg, B. (2017). A robust and novel regression based fuzzy time series algorithm for prediction of rice yield. 2017 International Conference on Intelligent Communication and Computational Techniques (ICCT).

[11] Raja, S. K. S., Rishi, R., Sundaresan, E., Srijit, V. (2017). Demand based crop recommender system for farmers. 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR).

[12] Dey, U. K., Masud, A. H., Uddin, M. N. (2017). Rice yield prediction model using data