Crop Yield Prediction using ArtificialIntelligence

Cotton Yield Prediction

Priyanka B G [1], Rashmi G [2], Sanjana B J [3], Vinitha K R [4], Priyanka P [5], Rajendra R Patil [6]
B. E Students[1234] (ECE) at GSSS Institute of Engineering & Technology for women, Mysore, Karnataka, India.
Asst. Prof.[5] Dept of ECE, GSSSIETW, Mysore, Karnataka, India.HOD.[6] Dept of ECE, GSSSIETW, Mysore, Karnataka, India.
Visvesvaraya Technology University, Belagavi, Karnataka, India .

***Abstract*—Cotton production is the main agricultural activity in India. More than 350.000 Indian families depend on cotton harvest. Since cotton rust disease was first reported in the country in 1983, these families have had to face severe consequences. Recently, machine learning approaches have built a dataset for monitoring cotton rust incidence that involves weather conditions and physic cotton properties. This background encouraged us to build a dataset for cotton rust detection in Colombian cottons through data mining process as Cross Industry Standard Process for Data Mining (CRISP-DM). In this paper we define a proper data to generate accurate models; once the dataset is built, this is tested using classifiers as: KNN and Linear regression Trees By analyzing all these issues and problems like weather, temperature and several factors, there is no proper solution and technologies to overcome the situation faced by us. In India there are several ways to increase the economic growth in the field of agriculture. There are multiple ways to increase and improve the cotton yield and the quality of the cottons. Data mining also useful for predicting the cotton yield production.**

INTRODUCTION

Six states with Karnataka in the lead are the major producers of cotton in the country. Karnataka with a production of 3.04 lakh tones from an area of 7.94 lakh hectares followed by Andhra Pradesh, Maharashtra, Bihar, Orissa and Tamil Nadu are major cotton producing states of India. In India, Cotton cultivation occupies about 1.48 M Ha area with average yield

0.6 MT/acre. Cotton production follows a systemic weather risk as about 80 per cent of the area is under rain-fed production. In terms of productivity, Bihar leads with 1402 kg/ha followed by Tamil Nadu with 1328.7 kg, although both the states have less than 25000 hectares under the cotton which is mostly irrigated. The average productivity at all India level was 900 kg/ha depending on the climatic conditions and irrigation, which are critical factors for high yields.

This work talks about K-Nearest Neighbor Algorithm and Linear Regression this algorithm does not have any learning phase, because every time a classification is performed it uses a training set. The assumption behind the k-nearest neighbor algorithm is that a similar classification is produced by similar samples. The similar known samples used for assigning a classification to an unknown sample are described by the parameter K.

The comparison for the results obtained by KNN Algorithm is given using Linear Regression Algorithm. Linear Regression is a linear approach for modelling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variables) denoted x.

I. OBJECTIVE

- "The cotton yield prediction" helps to predict the yield based on the consideration of different types of attributes like Temperature, Rainfall, Soil ph. level and Nitrogen.
- "The cotton yield prediction" predicts yield based on individual farmers past yield records of particular plot.
- "The cotton yield prediction" predicts the yield using "KNN Algorithm" and a comparison study has been given based on "Linear Regression".
- "The cotton yield prediction" is implemented using JAVA technology.

II. PROBLEM STATEMENT

"Cotton Yield prediction" has become a global issue and is an area of concern. All farmers will not be having much knowledge regarding how much yield can be produced in their plot with certain soil and weather condition.

A. Existing System

Nowadays, there are many agricultural sectors to provide guidance to the farmers regarding the cotton yield production. These agricultural sectors consider the Soil pH, Nitrogen and other fertilizers to predict the cotton yield and prescribe the measures to be taken to increase the cotton yield. Most of the time the weather conditions like Temperature ad Rainfall are not considered to predict cotton yield and hence there will be less accuracy in the predicted results as all there prediction is

done manually. There is no automation for the prediction of "cotton Yield".

III. SCOPE

- "Cotton Yield prediction" is an agricultural sectorapplication.
- "Cotton Yield prediction" area of concern in predicting the Cotton yield.
- "Cotton Yield prediction" is automation for Cottonyield prediction.

IV. METHODOLOGY

The system "Cotton Yield prediction" works on data mining techniques. The data mining techniques used here is KNN Algorithm and a comparison for the results obtained by KNN Algorithm is given by the Linear Regression Algorithm. Data mining comes up with a set of tools and techniques which when applied to this processed data, provides knowledge to farmers for making appropriate decisions and enhance the production of yield of Cotton. By the consideration of the previous year data regarding the Rainfall, Temperature, Soil pH, Nitrogen and Yield of a particular plot of the farmer, how much yield can be produced is predicted using the KNN Algorithm. "Cotton Yield prediction" has become a global issue and is an area of concern. It a condition where future prediction about the cotton yield can be done based on the previous data collected and measures can be taken to improve the cotton yield production of Cotton for that particular season.

V. PROPOSED SYSTEM

Prediction of cotton yield has become a major issue in agricultural field and is an area of concern. This prediction gives a brief knowledge to the farmers regarding how much yield can be expected on their plot based on the analysis of previous year records. This prediction is done using the Classification Algorithm. Proposed system is automation for Cotton Yield prediction using the classification technique "KNN Algorithm". A comparison for the result obtained by the KNN Algorithm is given using Linear Regression.

VI. HARDWARE AND SOFTWARE REQUIREMENT

Software Requirements:
Front End: HTML, CSS,
BootstrapBack end : MySQL
Tool : JSP, Servlets, Ajax,
JSonID : NetBeans

Hardware Requirements:
Ram : 512
MbHard Disk :
50 GB

VII. DATASETS

In this work, the datasets have been analyzed based on different types of parameters which are inter-related with each other and these parameters affect in a major way for the prediction of the Cotton yield. The parameters which are majorly considered in this work are Average Rainfall, Average Temperature, Average Nitrogen, Average Soil Ph value and the previous year's Yield data of the same plot. These parameter values are analyzed first and then related which each other based on the maximum and minimum value required for the Cotton yield production.

A. The data obtained by the farmer

- Name of the Taluk
- Type of cotton– Cotton
- Plot's Survey Number
- Season – Summer / Kharif
- Soil Ph value
- Nitrogen value
- Yield obtained for particular year (Approximately forpast 7 to 8 years).

B. The data obtained by the website and other resourses

- Average Rainfall data of every month from 2012 to 2019 (for all the seven taluks)
- Average Temperature data of every month from 2012to 2019 (for all the seven taluks)

C. Data sets collected

Sl No	Parameters	No of Datasets
1	Target Report	11,340
2	Average Rainfall	700
3	Average Temperature	700
4	Survey datasets obtained from Farmers regarding - Soil ph, Nitrogen and previous year's yield data	50

VIII. LITERATURE SURVEY

[1] Prediction of Cotton Yield using Regression Technique (International Journal of Soft Computing, Vol. 2, Issue 9, 2017)Author: Adithya Shastry, H A Sanjay and E Bhanusree

Summary: In this paper, author focused on the development of Regression techniques on Cotton field. This paper involves predicting the yield of cottons from available historic data like whether parameters, soil parameters and historic cotton yield. Regression is a data mining function that predicts a number. Different Regression techniques such as quadratic, pure- quadratic, interactions and polynomial are used for predicting the yields of cotton.

[2] Evaluation of Modified K-Means Clustering Algorithmin Cotton Prediction
International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume-4 Number-3 Issue-16 September-2014 **Author:** Utkarsha P. Narkhede, K.P.Adhiya
Summary: This paper presents the evaluating the Modified K-Means Clustering Algorithm in Cotton. Clustering is a data mining algorithm and plays significant role for extracting knowledge and update of information. This paper concentrating on Clustering technique applied in cotton dataset has resulted in novel approach which has significance success in predicting cotton. The drawbacks are overcome by proposing modified k-Means clustering algorithm which used the formulated value to initialize cluster centers and to determine number of clusters. This paper demonstrates about modified k- Means clustering in cotton prediction by increasing quality and accuracy count.

[3] Modified Naïve Bayes Based Prediction Modeling for Cotton Yield Prediction
World Academy of Science, Engineering and Technology International Journal of Biological, Biomolecular, Agricultural, Food and Biotechnological Engineering Vol:8, No:1, 2014 **Author:** Kefaya Qaddoum
Summary: This paper works on issues of irrelevant predictors and redundant predictors for the naive Bayes model. The utilized model that, initially inspired by the naive Bayes scheme, deals reasonably well with these false predictors. This has been proved empirically on several data sets, where different numbers of irrelevant and redundant predictors have been added. This paper proposed a method had been found much better than naive Bayes model, since the L1-penalty deals with redundancy then the redundant predictors could be discard.

[4] "Analysis of Soil Behavior and Prediction of Cotton Yield Using Data Mining Approach". 2015 International Conference on Computational Intelligence and Communication Networks. **Author:** Monali Paul, Santosh K, Vishvakarma and Ashok Verma
Summary: This work presents a system, which uses data mining techniques in order to predict the category of the analyzed soil datasets. The category, thus predicted will indicate the yielding of cottons. The problem of predicting the cotton yield is formalized as a classification rule, where Naive Bayes and K-Nearest Neighbor methods are used.

[5] "A predictive modeling approach for improving cotton cotton productivity using data mining techniques" Turkish Journal of Electrical Engineering & Computer Sciences **Author:** Anitha Arumugam
Summary: The proposed research aims to develop a predictive model that provides a cultivation plan for farmers to get high yield of cotton cottons using data mining techniques. Unlike statistical approaches, data mining techniques extract hidden knowledge through data analysis. The data set used in this research for mining process is real data collected from farmers cultivating cotton along the Thamirabarani river basin. K-means clustering and various decision tree classifiers are applied to meteorological and agronomic data for the cotton cotton. The performance of various classifiers is validated and compared. Based on experimentation and evaluation, it has been concluded that the random forest classifier outperforms the other classification methods. Moreover, classification of clustered data provides good classification accuracy.

IX. SOFTWARE REQIREMENT SPECIFICATIONS

The presentation of the Software Requirements Specification (SRS) gives a review of the whole SRS with reason, scope, definitions, abbreviations, contractions, references and diagram of the SRS. The point of this report is to assemble, dissect, and give a top to bottom knowledge of the total "Cotton yield prediction" by characterizing the difficult articulation in detail. The point by point necessities of the Cotton yield prediction purchasing conduct – client related capacities are given in this archive.

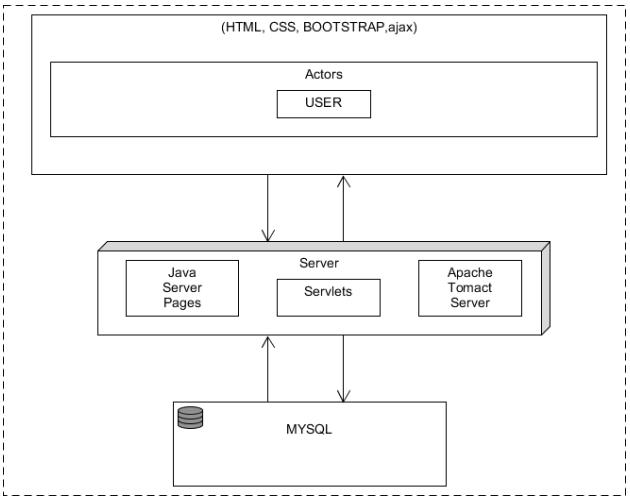
A. Purpose

The Purpose of the Software Requirements Specification is to give the specialized, Functional and non-useful highlights, needed to build up a web application App. The whole application intended to give client adaptability to finding the briefest as well as efficient way. To put it plainly, the motivation behind this SRS record is to give an itemized outline of our product item, its boundaries and objectives. This archive depicts the task's intended interest group and its UI, equipment and programming prerequisites. It characterizes how our customer, group and crowd see the item and its usefulness.

B. Scope

The Scope of this framework is to presents a survey on information digging strategies utilized for the expectation of Cotton yield prediction. It is obvious from the framework that information mining strategy, similar to grouping, is profoundly productive in expectation of Cotton yield prediction.

C. Software Architecture



D. Acronyms and Abbreviation

1) Front End: JSP

JavaServer Pages (JSP) is an innovation that encourages engineers to make progressively produced pages dependent on HTML, XML, or other report types. Delivered in 1999 by Sun Microsystems, JSP is like PHP and React's JSX, yet it utilizes the Java programming language.

JavaServer Pages is an innovation for creating WebPages that underpins dynamic substance. It assists designers with embeddings java code in HTML pages by utilizing uncommon JSP labels, the greater part of which start with <% and end with %>.

2) Servlets

Java Servlets are server-side Java program modules that process and answer customer demands and actualize the servlet interface. It helps in upgrading Web server usefulness with negligible overhead, upkeep, and backing. A servlet goes about as a middle person between the customer and the server. As servlet modules run on the server, they can get and react to demands made by the customer. Solicitation and reaction objects of the servlet offer an advantageous method to deal with HTTP asks for and send text information back to the customer. Since a servlet is incorporated with the Java language, it additionally has all the Java highlights, for example, high convenience, stage freedom, security, and Java information base network.

3) Java

Java is a universally useful, elevated level programming language created by Sun Microsystem. Java is inexactly founded on C++ grammar and is intended to be Object-Oriented. The java compiler changes over source code into Byte Code, which is secure and convenient across various stages. These byte codes are fundamental guidelines embodied in a solitary sort, to what in particular is known as java virtual machine (JVM), which dwells in a standard program. JVM is accessible for practically all OS. JVM changes over these byte codes into machine-explicit directions at runtime.

4) HTML

A long time back, HTML was an abbreviation utilized absolutely by senior programming designers and technically knowledgeable undergrads. As the world has gotten more digitalized, the interest for HTML coding abilities has extraordinarily expanded. HTML may seem like an unfamiliar term to a few. It represents Hypertext Markup Language. Hypertext Markup Language is the language that is utilized on sites to show textual styles, hues, connections, and pictures to the site guests. Inside the language, there are a few labels that are utilized to empower the site designer to alter the manner in which the site looks. These labels and codes are then perused by Internet programs, which show the site as indicated by the particulars recorded in the HTML codes. Guests to the site at that point see the individual pages inside the site as the designer planned them, without seeing the rundown of codes and labels.

5) Backend: mysql

MySQL is the most well known Open Source Relational SQL information base administration framework. MySQL is extraordinary compared to other RDBMS being utilized for creating online programming applications.

MySql is an incredible information base. It's generally excellent and complimentary. Numerous engineers on the planet chose MySql for building up their site.

X. SOFTWARE DESIGN

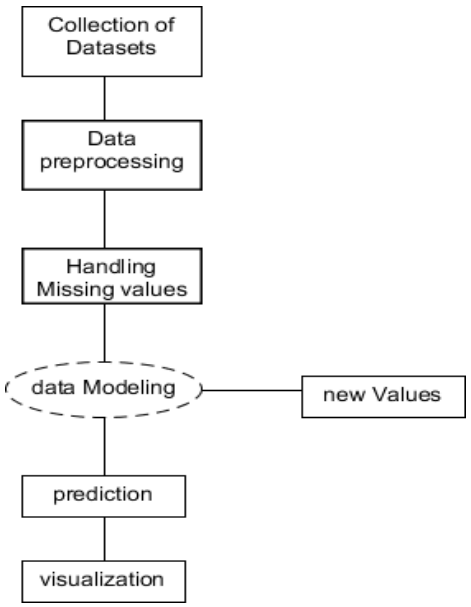
A. Introduction

The Software Design will be used to aid in software development for android application by providing the details for how the application should built. Within the Software Design, specifications are narrative and graphical documentation of the software design for the project includes use case models, sequence diagrams and other supporting requirement information.

B. Scope

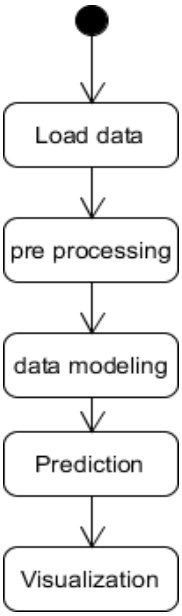
This software Design Document is for a base level system, which will work as a proof of concept for the use of building a system that provides a base level of functionality to show feasibility for large-scale production use. The software Design Document, the focus placed on generation of the documents and modification of the documents. The system will used in conjunction with other pre-existing systems and will consist largely of a document interaction faced that abstracts document interactions and handling of the document objects. This Document provides the Design specifications of cotton.

C. Architectural Diagram



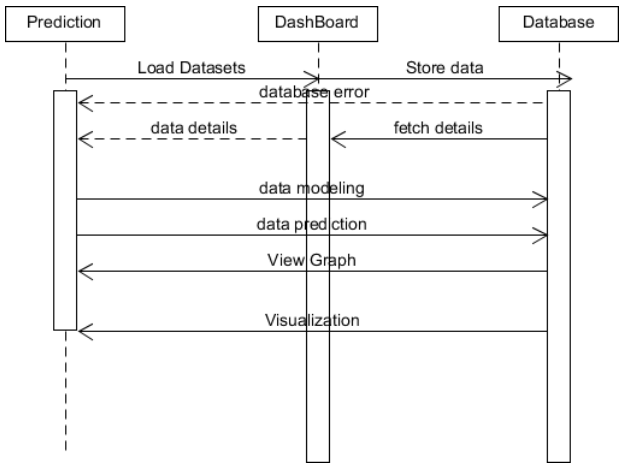
1) Activity Diagram

An activity diagram outwardly presents a progression of activities or stream of control in a framework like a flowchart or an information stream chart. Action graphs are regularly utilized in business measure demonstrating. They can likewise depict the means in a utilization case chart. Exercises demonstrated can be consecutive and simultaneous. In the two cases, an action outline will have a start (an underlying state) and an end (a last state).

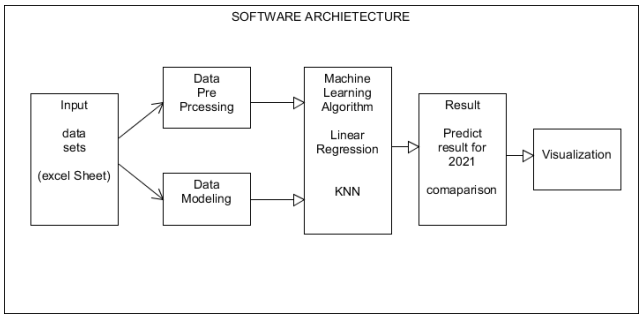


2) Sequence Diagram

Sequence diagram depict cooperates among classes as far as a trade of messages after some time. They're likewise called occasion charts. A grouping chart is a decent method to envision and approve different runtime situations. These can assist with anticipating how a framework will act and to find duties a class may need to have during the time spent demonstrating another framework.



D. Software Architecture diagram



System architecture is a conceptual model that defines the structure, behavior and more views of a system. A system architecture can comprise system components, the expand systems developed, that will work.

XI. IMPLEMENTATION

The project is implemented using java which is an object oriented programming language. This project is implemented using java programming language. Both servlet and JSP technologies are used to create a web application. Servlet are java programs are precompiled which can create dynamic web contents. There are many interfaces and class in the servlet API such as Http servlet, servlet request, servlet response etc. JSP is used to create a web application just as servlet.it can be thought of as a extension to servlet because it provides more functionality than servlet. MySQL server is used as a backend.

A. Algorithms

The Algorithms are the process or set of rules to be followed in calculations or other problem solving operations and are also used in the prediction process.

1) K- Nearest neighbour(KNN) algorithm

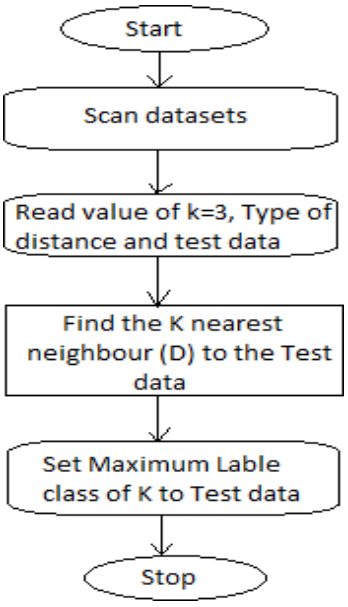
The K-nearest neighbor algorithm is a method for classifying objects based on closest training examples in the feature space. With previously labeled samples as the training set S, the KNN algorithm constructs a local sub region $R(x) \subseteq \Re$ d of the input space, which is situated at the estimation point x. The predicting region R(x) contains the closest training points to x, which is written as follows:

$R(x) = \{x^ \mid D(x, x^) \leq d(k)\}$,

For a given observation x, the decision g(x) is formulated byevaluating the values of k[y] and selecting the class that hasthe highest k[y] value

$g(x)=\{1,-1,k[y=1] \geq k[y=-1], k[y=-1] \geq k[y=1]$.

Thus, the decision that maximizes the associated posterior probability is employed in the KNN algorithm.



2) Linear Regression Algorithm

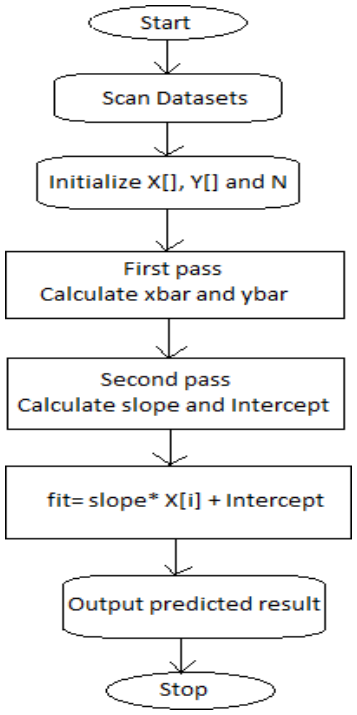
Linear regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple Linear Regression. In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models. Most commonly, the conditional mean of the response given the values of the explanatory variables (or predictors) is assumed to be an affine function of those values; less commonly, the conditional median or some other quintile is used.

Where T denotes the transpose, so that xiTβ is the inner product between vectors xi and β.

Often these n equations are stacked together and written in matrix notation aswhere

•Y is a vector of observed values of the variable called the regressand, endogenous variable, response variable, measured variable, criterion variable, or dependent variable

•X may be seen as a matrix of row-vectors Xi, or of n-dimensional column-vectors Xj.



XII. FUTURE SCOPE AND CONCLUSION

A. Conclusion

This project is an agricultural sector application which helps the farmers in the predicting the cotton yield based on the previous year datasets. Famers can check the predicted cotton yield for their plot by entering the past data of their plot. It is automation for cotton yield prediction and is an efficient and is economically faster.

It is successfully accomplished by applying the KNN Algorithm for cotton yield prediction and Linear Regression Algorithm for giving a comparison result for KNN algorithm. The Classification technique comes under data mining technology. These algorithms take the previous datasets as input and predict the cotton yield based on the previous datasets.

B. Scope for future work

•The present system is developed for the prediction of Cotton only, in future a system can be developed to predict cotton yield for different types of cottons, vegetables, flowers and so on.

•The present system is developed for the seven different taluks of Prakasam only; future enhancement can be made by developing a system where prediction can be done for different cities and their taluks.

•The present system outputs the result based on KNN Algorithm and Linear Regression, whereas a system can be developed by the fusion of these algorithms as well as a comparison for KNN algorithm.

XIII. REFERENCE

[1] D Ramesh, B Vishnu Vardhan. “Data Mining Techniques and Applications to Agricultural Yield Data”. International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 9, September 2013.

[2] Ami Mistry and Vinita Shah.“Brief Survey Of Data Mining Techniques Applied To Applications Of Tobbaco”. International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 2, February 2016.

[3] A.T.M Shakil Ahamed, Navid Tanzeem Mahmood and Nazmul Hossain. “Applying Data Mining Techniques To Predict Annual Yield Of Major Cottons And Recommend Planting Different Cottons In Different Districts In Bangladesh”. Department of Electrical and Computer Engineering, North South University, Bangladesh.

[4] Monali Paul, Santosh K, Vishvakarma and Ashok Verma. “Analysis of Soil Behavior and Prediction of Cotton Yield Using Data Mining Approach”. 2015 International Conference on Computational Intelligence and Communication Networks.

[5] Datasets from “Karnataka State Natural Disaster Monitoring Center”

[6] Datasets from “Directorate of Economics and Statistics” ANNUAL RAINFALL REPORT OF 2010, 2011, 2013, 2014, 2015.

[7] Soil datasets from “Rashtriya Chemical and Fertilizers Ltd” survey, Suttur.

[8] Soil datasets from “Soil, water and cotton testing center”, Suttur.