IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol-6, Special Issue-8, Oct 2018 E-ISSN: 2347-2693

CROP YIELD PREDICTION AND SOIL DATA ANALYSIS USING DATA MINING TECHNIQUES IN KRISHNAGIRI DISTRICT

[1*]K. Samundeeswari, [2]K. Srinivasan

[1,2]Govt. Arts College for Women, Periyar University Constituent College Of Arts &Science, Tamilnadu, India

Available online at: www.ijcseonline.org

Abstract- The objective of this work is to explore the soil data analysis for crop yield prediction in KrishnagiriDistrict by comparing various data mining techniques which gives the maximum accuracy. Analyzing soil provides major contribution to support the farmers [2]. In this paper one of major parameter which is used to increase crop production is considered – soil, and also explores various proposed algorithms for analyzing soil using data mining techniques and different data mining algorithms are applied to soil data set to predict its soil fertility.

Keywords: Crop Yield,Soil Data, Agricultural Yield Prediction, K-Means, Support Vector Machine (SVM), Multiple Linear Regression (MLR)

I. INTRODUCTION

Data mining is the techniques used to extract the hidden predictive information from huge databases. Very important issue in agricultural problem is yield prediction.In agriculture, productivity of crop is one of the main issues. The farmer is interested to know how much yield is about to expect. In earlier days, the crop yield prediction was carried out only by the experience of the farmers. As technology improved there are many ways that leads to know the productivity of crop[4]. The application of data mining technique is one of the best ways to provide the solution for this problem.

Crop yield predict analysis requires a model of crops, how it respond to soil factors. As predictions from different soils types, understanding the sources of soil texture, physical and chemical properties. Fertilizing and harvesting causes, impact on thosefactors. This study analyzes the various factors of soil in different areas of Krishnagiri district and find, which type of soil is more suitable for the selected crop growth.

a. Soil Distribution in Krishnagiri District

Dataset Collection is part of work which are collected from Indian Council of Agricultural Research (ICAR)Krishnagiridistrict. Primary data for the soil types are acquired by field sampling [12]. These samples are sent to chemical and physical analysis at the soil testing laboratories; hence the dataset was collected from the agriculture department ofKrishnagiri district. It contains information about number of soil samples taken from various region of Krishnagiri district.

b. Administrative Details of Krishnagiri district

Administrative Details of Krishnagiridistrict,having administrative divisions of 5 taluks, 10 blocks and 626 villages.

Table: 1(Source: Survey Dept., Office of District Collector)

S. No.	Taluk	No. of villages	S. N.	Blocks	No. of villages
1	Denkanikottai	93	1	Thally	53
			2	Kelamangalam	35
2	Hosur	166	3	Hosur	79
			4	Shoolagiri	87
3	Krishnagiri	142	5	Krishnagiri	
			6	Veppanapalli	68
			7	Kaveripattanam	33
4	Pacchampalli	40	8	Bargur	39
5	Uthangarai	185	9	Uthangarai	145
			10	Mathur	**57**
Total		**626**			**626**

'Krishna' refers to 'black' and 'giri' refers to 'hill'. This district is gifted with black granite hillocks and named as "Krishnagiri". Krishnagiri district has prehistoric importance. Krishnagiri District came into existence from 9th February 2004, consisting of Hosur & KrishnagiriDivisions [13].

Krishnagiri is one of the districts of Tamil Nadu with natural resources has 2,040 Sq. Kms of forest cover and it is its unique feature. The hill ranges of this district are called by the name 'Melagiri'.

The major types of forest seen here are tropical, deciduous forests, thorny shrubs and bamboo forest. Dense forest cover is there in Denkanikottai region. The other region contains shrubs, hills and hillocks with bushes.

Krishnagiri District has 2 Municipalities, 10 Panchayat Unions, 7 Town Panchayats, 352 Village Panchayats and 636 Revenue Villages.

c. Geography of Krishnagiri District

Krishnagiri district is located approximately between 11'12N and 12'49N of the North Latitude and between 77'27E and 78'38E of east longitude. The total geographical area of the district is 5143 Sq. Km. Krishnagiri district is elevated from 300m to 1400m above the mean sea level [13].

The district is surrounded by Vellore and Tiruvannamalai districts in the East, Karnataka in the West, Andhra Pradesh in the North, Dharmapuri district in the South.The region came under the rule of Krishna Deva Raya and hence it might have been named after this king.

d. Soil Types of Krishnagiri District

a. Soil Types:

Soils have been classified into Black soil, mixed soil, red loamy soil, gravelly and sandy soils. Grayish brown, Sandy loam and clay loam are the soil types found in the Krishnagiri.Red loamy and sandy soils are predominant in HosurTaluk. Vast stretches of loam soils and black soils occur in Krishnagiri district[12].

b. Soil classification

Soil is a vital component in all ecosystem, in fact, our very existence depend on the 6-12 inches underneath our feet. Many soil samples are taken from various regions of Krishnagiri districts and these samples are analyzed to find its physical properties, its acidic and alkalinity. From this analysis the soils are classified based on their texture, color, drainage and its alkalinity. The following table lists out the soils types of various soils presented in various places of Krishnagiri district.

Table 2: Soil Types of Krishnagiri District

Sl. No.	Taluk	Soil Type
1	Uthangarai	Dark reddish brown, Sandy clay loam
2	Pochampalli	Dark reddish brown, Sandy loam
3	Krishnagiri	Grayish brown, Sandy loam and clay loam
4	Hosur	Dark reddish brown, Sandy loam
5	Denkanikottai	Dark reddish brown, sandy and clay loam

c. Irrigation Pattern :

In Krishnagiri district wells and canals are the major sources of irrigation contributing sources of irrigation contributing to 81.19 and 10.71 per cent of the total area irrigated of the district respectively. Tanks accounted for less than five per cent.

Table 3:Irrigation Pattern of Krishnagiri District

Revenue Divisions :	Revenue Taluks :	Panchayat Unions :	
Krishnagiri	Krishnagiri	Kelamangalam	Krishnagiri
Hosur	Hosur	Thally	Kavaripattinam
	Pochampalli	Hosur	Bargur
	Uthangarai	Shoolagiri	Mathur
	Denkanikottai	Vepanapalli	Uthangarai

Rain Fall:

Table 4: Rainfall of Krishnagiri District

Season	Rain fall (mm)
Winter	:19.00
Summer	:311.60
South West monsoon	:160.20
North East Monsoon	:239.70
Grand Total	**:830.50**

d. Land Use Pattern

The total geographical area of the district is 5.14 lakh hectare, of which net sown area accounts for 35.03 per cent[13]. The area under forest accounted for 39.35 per cent. Fallow and current fallow lands account for less than 5 per cent in the district.

II.THE IMPORTANCE OF CLIMATE

Climate has Macroclimate and Microclimate.

Macroclimate refers to large-scale regional weather conditions and long term trends in cloud over, humidity and precipitation, and temperature.

Microclimate is a function of short term, local fluctuations in light, temperature, and moisture at a particular site. These factors are affected by distinctive site conditions such as

wind barriers and shade created by trees and other obstacles, the aspect (direction) and slope of a hillside, frost pockets, streams banks, etc.Both macroclimate and microclimate contribute to growing conditions, so knowledge of these factors should inform your selection and placement of crops [1].

III. STUDY ABOUT SOIL

a. **Soil**

Soil is a living breathing,natural entity composed of solids, liquids and gases.

Soil has five major functions

1. Provides a habitat for organisms
2. Recycles waste products
3. Filters water
4. Serves as an engineering material
5. Provides a medium for plant growth

b. **Types of Soil**

There are basically three types of soil which are sand, silt and clay. But most soils aren't pure they are composed of a combination of sand, silt and clay

Sand

Weathered rock fragments

The first type is sand which actually consists of small particles of **weathered rock**. Sandy soil are one of the worst types of soil to grow any kind of plants, this soil has largest particles that prevents from retaining water, making it hard for the roots of plants to establish.But it is good for drainage.

Silt

Silt is known to have much smaller particles than sandy soil so it's smooth to touch and quite fine, holding water better than sand [2]. And if water is added to it, it will hold water well and feels slick and smooth.

Clay

Clay, it happens to contain the smallest particles amongst the three, placed tightly together with each other with little or no air space and it has good water storage qualities, while at the same time making hard for moisture and air to penetrate it. It's sticky to touch when wet, but smooth when dry. Clay is the densest and heaviest type of soil. Also, clay does not drain well or provide space for plant roots to flourish.

Loam

Loam is the fourth types of soil. Even though it is a combination of sand, silt and clay. It is the gardener's favorite kind of soil. It contains a balance of all three soil materials being silt, sand,clay and also happens to have humus. It also has a higher pH and calcium levelsbecause of its previous organic matter content.

c. The Components of Soil

Managing and Understanding our soil is a very important step in ensuring the long-term sustainability of agriculture. It is important to understand the soil health issues that are interrelated and difficult to treat as single issues[3]. With this, it is important to manage soils in relation with their physical, chemical and biological properties.

Table5: Examples of physical, chemical and biological soil features are

Physical features	Chemical features	Biological features
Texture	pH	Organic matter
Structure	Salinity	Macro fauna (worms, arthropods)
Water infiltration & drainage	Sodicity	Soil bacteria. e.g. rhizobia
Topsoil structure	Total nitrogen	Fungal hyphae
Soil colour	Total phosphorus	Micro fauna (nematodes, protozoa)
Slaking	Trace elements	Abundance of roots

The physical, chemical and biological features of soil interact and need to be managed .

For example,

Soil's with a low pH (less than pH 5) result in chemical imbalances

- Such as aluminum toxicity and deficiencies of phosphorus
- Trace elements such as calcium and molybdenum.

Very low pH (less than pH 4) leads to soil physical breakdown

- The clay structure of the soil is broken down.
- Acid soils also impact on soil biota, reducing earthworm numbers
- Making Rhizobia less effective.

d. **Physical Properties of soil**

The physical characteristics of a soil can be seen, felt, or measured. These include color, texture, structure, and water-holding capacity. Such properties determine the suitability of soil as a growth medium. Some physical properties, such as texture, are not economically feasible to change on a large scale.

A soil's fertility, which is a chemical property, is easier to change than the soil's physical properties.

A. Color

Organic matter, soil minerals present, and the drainage conditions all influence soil color. Color alone does not indicatethe soil quality, but color does provide clues to be used for certain conditions.

For example,

- Light or pale colors in grainy topsoil are frequently associated with low organic matter content, high sand content, and excessive leaching.
- Dark soil colors may result from poor drainage or high organic matter content.

ii. Texture

Soil texture, which refers to the proportions of sand, silt, and clay, influences nearly every aspect of soil use and management.

The following table to assess soil:

Table 6: Soil Texture Properties

Texture class	Properties of moist soil ball
Sands	Won't form a ball. Forms a ribbon less than 10 mm. Feels very sandy and not sticky at all. Clay content 0 to 10%.
Sandy loams	Able to form a ball. Forms a ribbon 15-25 mm long. Feels sandy and slightly sticky. Clay content 10 to 20%.
Loams	Forms a smooth ball, ribbons to 25-40 mm. Feels slightly sandy and moderately sticky. Clay content 20 to30%.
Clay loams	Forms a smooth, plastic ball, ribbons to 40-50 mm. Almost no sandy feel. Distinctly sticky. Clay content 30 to35%.
Light clays	Forms a smooth, plastic ball, ribboning to 50-75 mm. Very sticky. Clay content 35 to 45%.
Medium to heavy clays	Forms a smooth, extremely plastic ball. Ribbon more than 75 mm. Feels very sticky with no sand. It is more difficult to mould than light clay. Clay content greater than 45%.

Table 7: Particle type, number of particles per gram, and the average surfacearea per gram.

Particle Type	Diameter (mm)	Number of Particles per gram
Clay	< 0.002	90,260,853,000
Coarse sand	0.50 -1.00	720
Fine sand	0.10 - 0.25	46,000
Medium sand	0.25 - 0.50	5,700
Silt	0.002 - 0.05	5,776,000
Very coarse sand	1.00 - 2.00	90

B. Chemical Properties of Soil

There are strong relationships between soil physical properties and soil chemical properties. Soil with high CEC not only holds more nutrients, they are able to buffer or avoid rapid changes in the soil solution levels of these nutrients [7]. A soil test will tell you the CEC number of your soil. Soil high in clay, silt, or organic matter will have a CEC number of 10 or greater, and no remediation is needed. Sandy soil will have a CEC number between 1 and 5. Adding organic matter to these soils will help increase the CEC.

C. Soil structure and organic matter

- Organic matter is the remains of living things or products of living things in the soil. Organic matter is important for soil structure.
- Organic matter on the soil surface (such as wheat stubble residues) protects the surface from the action of raindrops, reducing surface compaction and hard setting. Organic matters bind sandy and silty soils together to improve water infiltration through the soil. Organic matter also acts as a buffer against the forces of compaction.
- Organic matter diminishes continuous cropping and cultivation the soil very quickly, leading to soil structural decline.

D. Soil Drainage

Soil drainage is how well the soil holds water, is determined by soil texture as well as other elements in the soil.

- **Water Holding Capacity**
- **Saturation**
- **Field capacity**
- **Permanent wilting point**
- **Available water**

Soil Compaction, compression of soil by heavy machinery or animal traffic, can decrease the soil pore space leading to decrease soil air and increase water and bulk density (harder soil). This can decrease the quality of the soil and make it a more stressful environment for plants-whether crops or forage-to live in.

E. Soil Nutrients

Plants pull the nutrients from the soil through their roots[9]. These nutrients are grouped into **Macronutrients** and **Micronutrients**.

Macronutrients are needed at high levels in the plants, and consist of the familiar Nitrogen, Phosphorus, and Potassium, usually abbreviated "N-P-K".

Micronutrients are just as essential as macronutrients, but in lesser quantities.

Macronutrients and their use in plants:

- ➢Carbon- Air- Sugars, fats, proteins, DNA, etc.
- ➢Oxygen- Air/Water- Sugars, fats, proteins, DNA, etc.
- ➢Hydrogen- Water- Sugars, fats, proteins, DNA, etc. pH regulation
- ➢Nitrogen- Soil- Proteins
- ➢Potassium- Soil- Ion balance in the cell
- ➢Phosphorus- Soil- Molecules used in all chemical reactions
- ➢Sulfur- Soil- Proteins, co-enzymes
- ➢Calcium- Soil- Cell wall
- ➢Magnesium- Soil- Chlorophyll, enzymes
- ➢Silicon- Cell Wall

Micronutrients and their use in plants:

- ➢Chlorine- Photosynthesis, ion balance in cells
- ➢Iron- Chlorophyll, enzymes
- ➢Boron- Carbohydrate metabolism
- ➢Manganese- enzyme function, root growth, and chlorophyll synthesis
- ➢Sodium- Photosynthesis, ion balance in cells
- ➢Zinc- Enzyme function
- ➢Copper- Enzyme function
- ➢Nickel- Enzyme function
- ➢Molybdenum- Nitrogen metabolism
- ➢Cobalt- Vitamin B12

Several factors improve a plant's ability to use nutrients:

- ➢Type of soil
- ➢Soil pH
- ➢Types of nutrients in the soil
- ➢Amount of soil water
- ➢Anything that affects the plant's growth

F. SOIL pH

Soil pH is a measure of the soil relative acidity or basicity. The pH scale ranges from 0 to 14. A pH of 7 is a neutral state, representing the value found in pure water [10]. Values above 7.0 are basic, while values below 7.0 are acidic. The pH scale is logarithmic, meaning each unit has a 10-fold increase of acidity or basicity. Thus, compared to a pH of 7.0, a pH of 6.0 is ten times more acidic, and a pH of 5.0 is 100 times more acidic.

G. Way to improve the Soil

Good aeration and drainage, as well as the ability to hold adequate moisture and nutrients, are key components of an ideal soil environment. some of the most important strategies for improving soil quality:

- ➢Minimize soil compaction (do not walk on garden beds or work wet soil)
- ➢Reduce drainage problems.
- ➢Decrease erosion.
- ➢Plant a cover crop
- ➢Incorporate organic matter.
- ➢Provide a 1- to 3-inch layer of organic mulch on the soil's surface.

Table 8:Organic materials and their application rates.

Organic Material	**Amount to be Added per 100 Square Feet**
Compost	10–20 cubic feet
Corncobs	50 pounds (2 bushels)
Hay	60 pounds (1 bale)
Leaves	75 pounds (3–4 bushels)
Sawdust	50 pounds (2 bushels)
Straw	60 pounds (1 bale)
Wood chips	50 pounds (2 bushels)

IV. PLANT NUTRITION AND FERTILIZATION

Peopledo not have a clear idea about plant nutrition and fertilization. Plant nutrition refers to the needs of the plant and how a plant uses the basic chemical elements[8]. Fertilization is the term used when these elements are supplied to the soil as amendments. Adding fertilizer during unfavorable growing conditions will not enhance plant growth and may actually harm or kill plants.

Table 9: Relative amounts (out of 100) of the essential nutrients required by most plants.

Primary Nutrients	
Carbon (C)	45
Oxygen (O)	45
Hydrogen (H)	6
Nitrogen (N)	1.5
Potassium (K)	1
Phosphorus (P)	0.2
Secondary Nutrients	
Calcium (Ca)	0.5
Magnesium (Mg)	0.2
Sulfur (S)	0.1
Micronutrients	
Iron (Fe)	0.01
Chlorine (Cl)	0.01

Manganese (Mn)	0.005
Boron (B)	0.002
Zinc (Zn)	0.002
Copper (Cu)	0.0006
Molybdenum (Mo)	0.00001
Amounts unknown for Nickel (Ni) and Cobalt (Co)	

V. SOIL TESTING

Soil testing gives valuable information about pH and plant-available nutrients. Test your every two to three years before planting thereafter. Inexpensive soil test kits are unreliable. To accurately determine the soil characteristics and the proper amount of lime and fertilizer to apply, contact the NC Department of Agriculture & Consumer Services (NCDA&CS). The accuracy of these reports, however, depends on the quality of the sample submitted.

Tips for Collecting a Good Soil Sample

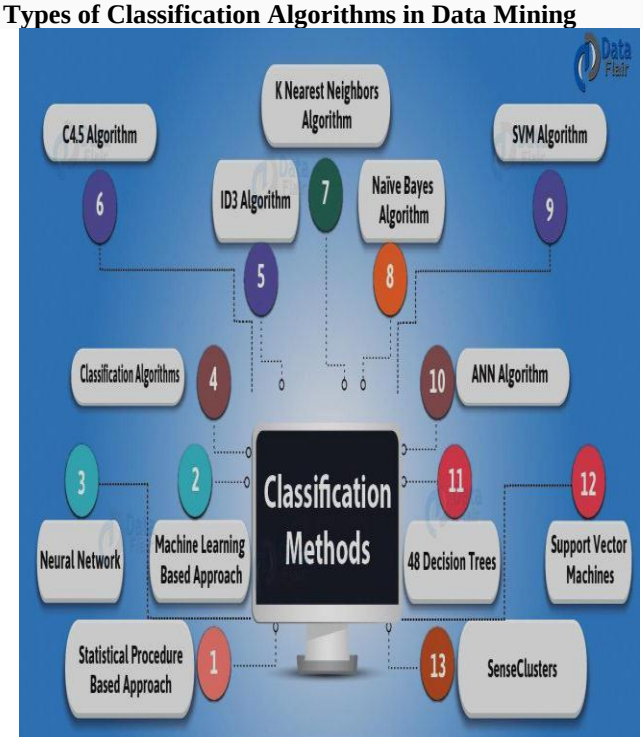
- Collect samples with stainless steel or chrome-plated tools. Using brass, bronze, or galvanized materials could contaminate the sample.
- The bucket in which material is collected should be made of plastic.
- Make sure the collection bucket is clean because even small amounts of residual lime or fertilizer can affect the test results.
- Avoid taking samples from areas that are obviously different from the norm, such as wet spots, compost piles, animal urine spots, and brush piles, or from under eaves or sites where trash has been burned.
- Remove large pieces of organic material, such as roots, stalks, and leaves, from the sample.
- For gardens, new lawns, and other cultivated areas, sample to the depth the soil has been, or will be, tilled[10]. For established lawns, collect the sample 2 to 4 inches deep. For trees and shrubs, take a sample to a depth of 6 inches near the plant's drip line.
- If using a trowel or spade, dig a hole, then take a slice of soil down one side. Repeat this procedure in five to eight spots for each area to be tested. Mix these cores together to obtain one composite sample. If the soil is very wet, it could be more difficult to mix, but do not attempt to heat the soil to dry it.
- Place about a pint of the composite sample for each area sampled in a soil testing box and label with a return address on the side of the box.
- Do not tape the boxes in any way. The lids are removed before the boxes go in the soil lab ovens, and tape makes this process difficult. Do not put the soil in a plastic bag before placing it in the box as doing so will prevent proper drying in the lab oven.

VI. CLASSIFICATION AND CLUSTERING ALGORITHM IN DATA MINING

Classification Algorithms

- Linear classifiers. Logistic regression. Naive Bayes classifier. Fisher's linear discriminant.
- Support vector machines. Least squares support vector machines.
- Quadratic classifiers.
- Kernel estimation. K-nearest neighbor.
- Decision trees. Random forests.
- Neural networks.
- Learning vector quantization.

Types of Classification Algorithms in Data Mining



Data Mining Classification Algorithms

These are the examples, where the data analysis task is Classification Algorithms-

- A bank loan officer wants to analyze the data in order to know which customer is risky or which are safe.
- A marketing manager at a company needs to analyze a customer with a given profile, who will buy a new computer.

Application of Classification Algorithms

- Email spam classification
- Bank customers' loan pay bank willingness prediction.
- Cancer tumor cells identification.
- Sentiment analysis.
- Drugs classification
- Facial key points detection
- Pedestrian's detection in an automotive car driving.

Clustering

Clustering is the process of making a group of abstract objects into classes of similar objects.

Clustering methods

Clustering methods can be classified into the following categories −

- Partitioning Method
- Hierarchical Method
- Density-based Method
- Grid-Based Method
- Model-Based Method
- Constraint-based Method

Applications of Cluster Analysis

- Clustering analysis is broadly used in many applications such as market research, pattern recognition, data analysis, and image processing.
- Clustering can also help marketers discover distinct groups in their customer base[8]. And they can characterize their customer groups based on the purchasing patterns.
- In the field of biology, it can be used to derive plant and animal taxonomies, categorize genes with similar functionalities and gain insight into structures inherent to populations.
- Clustering also helps in identification of areas of similar land use in an earth observation database. It also helps in the identification of groups of houses in a city according to house type, value, and geographic location.
- Clustering also helps in classifying documents on the web for information discovery.
- Clustering is also used in outlier detection applications such as detection of credit card fraud.
- As a data mining function, cluster analysis serves as a tool to gain insight into the distribution of data to observe characteristics of each cluster.

VII. CONCLUSION

Data mining is the emerging field of technology that can be used for the crop yield prediction. Various techniques of classification and clustering can be used to analyze the data. This paper focuses on analyzing the agricultural soil data to predict the yield of the crop with its soil types in soil fertilityatKishnagiri district using data mining techniques like K-Means, Support Vector Machine (SVM), and Multiple Linear Regression (MLR),and also to find the accuracy of data mining techniques by the following algorithms used in this paper for crop yield prediction.In future the work can be expanded and enhanced using the climatic factors, crop prediction, fertilization techniques based on seasons edict specific crop on various soils, the yield class attribute contains the value as Poor, Good, Very _Good, and Excellent depends on their yield value. Similarly the proposed study is toanalyzethe suitable crop for suitable soil, like paddy growth is better in red non-calcareous soil rather than the other soils,enhanced to find Black soil for cotton, Red Calcareous for Groundnut, Brown soil for sugarcane and Mixed soil for Maize using classification and clustering method in data mining.

REFERENCES

[1].N Hemageetha , G.M. Nasira,"Analysis of the Soil Data Using Classification Techniques for Agricultural Purpose"International Journal of Computer Sciences and Engineering,4(6)(2016).

[2].Huma Khan ,ShahistaNavaz, Dr. S. M. Ghosh3," A Survey on Various Data Mining Techniques in Field of Agriculture for Prediction of Crop Yield",International Journal of Science and Research (IJSR),6(5)(2017).

[3].E. Manjula,S. Djodiltachoumy,"A Model for Prediction of Crop Yield",International Journal of Computational Intelligence and Informatics, 6(4)(2017).

[4].P. Kanjana Devi.S.Shenbagavadivu" Enhanced Crop Yield Prediction and Soil Data Analysis Using Data Mining",International Journal of Modern Computer Science 4(6)(2016).

[5]. D Rames , B Vishnu Vardhan"Data Mining Techniques and Applications to Agricultural Yield Data" International Journal of Advanced Research in Computer and Communication 2(9)(2013).

[6].D Ramesh , B Vishnu Vardhan," Analysis Of Crop Yield Prediction Using Data Mining Techniques"International Journal of Research in Engineering and Technology(IJRET),4(011)2015.

[7].Ramesh A.Medar, "A survey on Data Mining Techniques for Crop Yield Prediction" –A research Article in ijarcsms ,2(9) 2014.

[8].G Ruß, "Data Mining of Agricultural Yield Data : A Comparison of Regression Models", Conference Proceedings, Advances in Data Mining – Applications and Theoretical Aspects, P Perner (Ed.), Lecture Notes in Artificial Intelligence 6171, Berlin, Heidelberg, Springer, 2009, pages : 24-37.

[9].Sellam,V., Poovammal, E., "Prediction of Crop Yield using Regression Analysis", Indian Journal of Science and Technology, Vol. 9(38), pp.1-5, 2016.

[10].Sujatha, R., Isakki, P., "A study on crop yield forecasting using classification techniques", International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE), pp.1-4, 2016.

[11].S. Veenadhari, Dr. Bharat Misra, Dr. CD Singh, "Data mining Techniques for Predicting Crop Productivity – A review Article", International Journal of Computer Science and Technology. (IJCST)2(1)(2011).

[12]. https://en.wikipedia.org/wiki/Krishnagiri_district

[13].https://krishnagiri.nic.in/about-district/history/

