Yield Prediction of Paddy based on Temperature and Rain Fall Using Data Mining Techniques

S.P. Aishwarya, Pramod Sunagar, Anita Kanavalli

Abstract--- The population of the world is estimated to be about 7 billion and might reach 9 billion in a few couple of years. With this growth in the global population, the world is facing challenges in food production. Prediction of agriculture is helpful in analyzing the risk, decide on storage, transportation and marketing. But rain and weather conditions are highly variable and, hence, it requires Data Mining. Data sets are extracted and analyzed to determine the rainfall patterns, humidity, wind speed and temperature, there by predicting the yield. The key idea here is to collect the data having various parameters affecting the yield, classify the data using KNN, then predict the yield using Apriori algorithm and analyze the productivity thus helping in decision making on marketing and risk. With the open source R Studio the graphical analysis is made. Since paddy is one of the basic food crop and also the major food crop grown in Karnataka state, the analysis is made on paddy yield considering two districts Koppal and Raichur which produces the major paddy yield.

Keywords--- Data Mining, KNN Classifier, Apriori Algorithm, R, Koppal, Raichur.

I. INTRODUCTION

The huge amount of data that is generated by the various applications may be structured, unstructured or semi structured and hence is coined as big data. The traditional file system is not efficient in storing, processing and analysing the generated data. Big data can be classified based on three main aspects they are Volume, Variety and velocity. Gartner introduced these elements in to big data.

Table I: Various Data Analysis Types

Analysis Type	Type of data	Architecture used	Category
Real-time	Data from sensors	Greenplum HANA	Parallel processing
Offline	Less response time	Scribe	Efficient data
Memory level	Memory for cluster	MongoDB	Real Time
Business intelligence level	Data more than memory level	Data analysis plan	Both offline and online
Massive level	Massive data	Map Reduce	Mostly Offline

Agriculture is known as back bone of India and is one of the major sectors of economy. There are many factors like rainfall, nutrients of soil, information about the seed, weather, pests etc. which can impact the yield and hence if we can provide a sensitive information about all these factors to the various stake holders an analysis can be made and finally given to the end user mainly farmers with which he can make effective decisions and there by improve the productivity. Data Mining is the process of analyzing, extracting and predicting the meaningful information from huge data to extract some pattern.

II. MATERIALS AND METHODS

A. *Existing System*

Prediction of agriculture is helpful in analyzing the risk, decide on storage, transportation and marketing. But rain and weather conditions are highly variable and, hence, it requires Data Mining. Data sets are extracted and analyzed to determine the rainfall patterns and temperature, there by predicting the yield. After the data is collected from the government data sets, it's preprocessed, cleaned and Predictive Apriori Algorithm with the data mining tool WEKA is used for analyzing the yield vs temperature and yield vs rainfall. The final analysis was that, in a vegetative phase as the temperature was increased even yield increased, but in reproductive stage with increase in temperature the yield decreases and with increase in rainfall the rate of growth is increasing and in maturation stage it increases with low temperature and low rainfall. Here only rainfall and temperature are considered, but there are many other factors such as soil pH, nutrients level, fertilizers which can be used etc. and better clustering algorithms which can analyses on various factors. [1]

There are three award winning projects of Government of India they are e-SAGU [25], BHOOMI [26], Sampark which have played an important role in bridging the gap between the agriculture technologies. a AQUA is another agro system developed by IIT Bombay which has helped the farmers in answering their queries. Krishimitra is an android based application for cotton crop prediction using RESTful services in Gujarat. Acropodia [28], Green phablet and Agrovoc are the projects developed by ICRISAT. Kisan Yojana, mKisan are few applications available. But this may be only specific to a particular region in India. [23]

Manuscript received September 16, 2019.

S.P. Aishwarya, Student, Dept. of CSE, MSRIT. (e-mail: aishwaryasprakash@gmail.com)

Pramod Sunagar, Assistant Professor, Dept. of CSE, MSRIT. (e-mail: pramods@msrit.edu)

Dr. Anita Kanavalli, Professor and Head, Dept. of CSE, MSRIT. (e-mail: anithak@msrit.edu)

B. Problem Statement

To implement a big data analytical framework which predicts the yield by analyzing the data collected so that the end user i.e. the farmer can make use of this prediction for decision making.

C. Proposed System

Apriori Algorithm which takes the input data having frequent transactions is used here where the input provided is the dataset having temperature, rainfall, humidity and wind speed. This input dataset is first labelled based on the classification data provided in to two districts that is Raichur and Koppal using KNN Classifier.

Then Apriori algorithm is used for predicting the yield. The result of Apriori and KNN can be compared to know the yield of the two districts. The data can be analyzed using the open source R Studio and the resulting analysis can be pushed to cloud and host an application which provides information about the yield. Similar grounds can also be applied for different crops of different districts.

D. The Study Area

The area of studies are Raichur and Koppal, the districts of Karnataka, in India, Raichur lies approximately 15.9268° N latitude, 76.6413° E longitude and Koppal lies at 15.3507° N latitude, 76.1554° E longitude. These districts are located in the northern maidan region of Karnataka. The climate in Raichur has the temperature of about 27.7 °C and near about 713mm of precipitation falls annually. The temperature in Koppal varies from 32°C to as high as 40°C. From July to October, the average rainfall is around 572 mm every year.Paddy is grown well on the alluvial soil or on the fertile river banks but.

This project is proposed for the paddy yield as it is one of the basic food crop and also the major food crop grown in Karnataka state, the analysis will be made on paddy yield considering two districts Koppal and Raichur which produces the major paddy yield.

E. Architecture

Apriori Algorithm is used to analyze the frequent item sets for data mining. It uses a level wise methodology to obtain the output where in it will scan the data in different levels to check the patterns and finally the prediction is made. For analysis of the results obtained from Apriori algorithm the open source R Studio can be used which supports various libraries and hence generating the graphs or histograms is easy and also is understandable for the end user.

Fig. 1: Architecture

F. Data and Sources

The data is collected from the open source sites which has normalized values. The system can also be tested against the actual data which can be obtained from the government. List of parameters of the Dataset is as below.

- Temperature
- Humidity
- Rainfall
- Wind Speed
- Yield

Table II: Description of Input Parameters

Variable	Unit	Classification
Temperature	°C	Numerical
Humidity	g/m3	Numerical
Rainfall	mm.	Numerical
Wind Speed	m/s	Numerical
Yield	Tons	Numerical

In the input data there are 8 columns for the different values given, which is represented below:

- ID: It defines the ID of each set of the data provided.
- Datatype: It represents which type of data is given. Here we are using normalized double datatype values, so we assign '1' in this field to represent the raw normalized values.
- Number of Parameters: It represents the number of parameters given for each input row. We assign 5 here, as there are five parameters.

G. Analytical Methods

The dataset including temperature, humidity, rainfall, wind-speed and yield were utilized in this study.

The algorithm of KNN was utilized to produce the relation and interesting correlation between the data sets. Euclidean distance is used to compute the nearest neighbors and based on the distance the data set is classified in to two classes that is Raichur and Koppal.

$$\textbf{Euclidean distance} = \sqrt[2]{\sum_{i=1}^{k}(x_i - y_i)^2}$$

The rules consuming the confidence and support, which are greater than the defined values are represented by algorithm as an output and lastly these rules are examined to generate the input data for the Apriori-Algorithm.

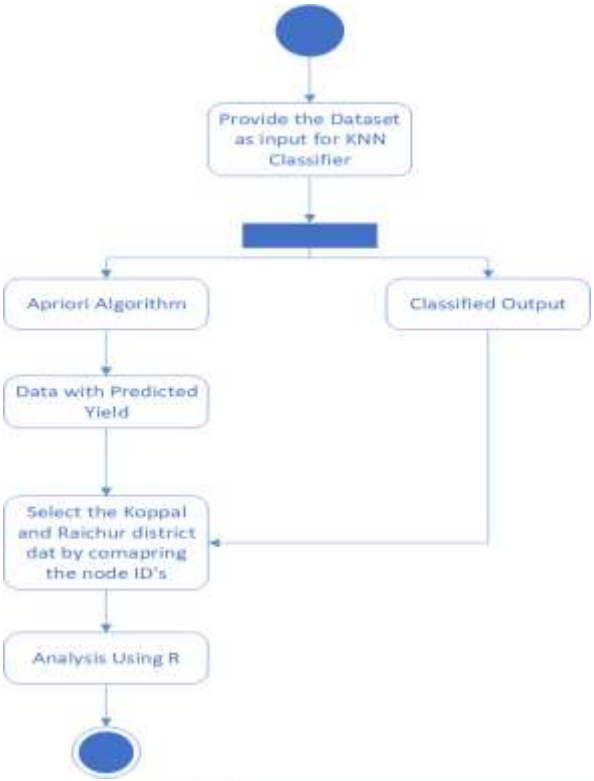

Fig.2: Activity Diagram

H. Methodology

In this project, the idea is to implement a computing system that predicts the yield of paddy for specific region by providing Temperature, Humidity, Rainfall, Wind speed as the input parameters along with the classification dataset. This system is based on Hadoop. Here firstly the Input file and Classification data file is copied from Local file system to Hadoop file system. Once the files are copied into the Hadoop file system, the KNN-classifier is allowed to classify the input datasets and once the classification of the data is done, execute the Apriori-Algorithm to predict the Yield of the paddy. The so obtained data is compared with the classified data obtained from the KNN-classifier to determine the nodes and distinguish between various regions or districts.

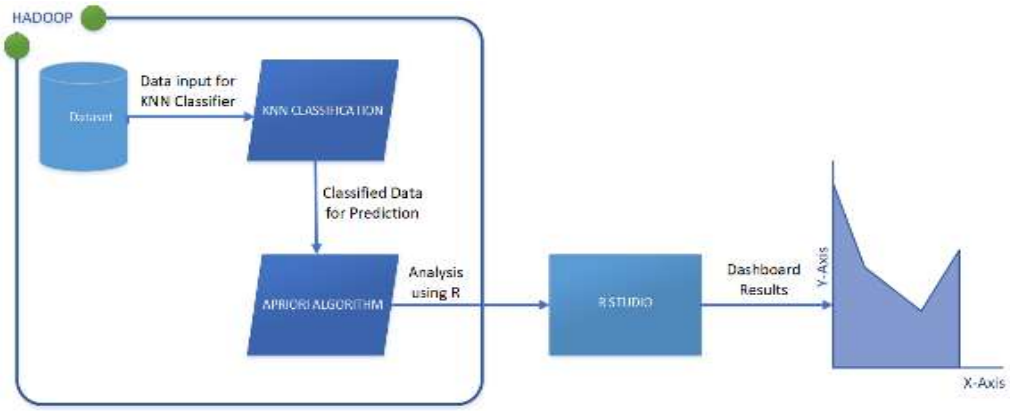

Fig.3: Methodology

III. RESULTS AND DISCUSSION

A. The Yield Prediction

The following Table-IV*:* represents the classified input data for Raichur and Koppal.

Table IV: Sample Output from KNN Classification

ID	Region
1	Raichur
2	Raichur
3	Raichur
4	Koppal
5	Koppal

After the above dataset is generated, it is given as input to the Apriori Algorithm.

By the end of execution of Apriori Algorithm, the following data sets as show in Table-V are generated, with the prediction of paddy.

Table V: Yield Generated from Apriori Algorithm

ID	Temperature	Humidity	Rainfall	Wind Speed	Yield
1	0.8357662	2.01335 49	1.12239 548	2.513194 14	5.34545 171
2	0.714321 58	2.03227 421	1.39594 125	2.901947 81	4.71723 209
3	1.052745 93	1.77666 873	1.25424 632	3.008867 77	5.25976 625
4	3.349677 44	3.57336 589	3.18458 978	0.607788 43	1.66117 607
5	2.850138 11	3.76512 611	3.04565 368	0.594705 15	2.35750 309

Once the above data is obtained, the datasets obtained from Apriori Algorithm and KNN-classifier is compared to retrieve the data for only Raichur and Koppal. That is the node ID's are compared. The following Table-VI and Table-VII are obtained after comparing the Table-IVand Table-V.

Table IV: Raichur Yield Prediction

ID	Region	Actual Yield	Predicted Yield
1	Raichur	5.25506338	5.34545171
2	Raichur	4.619899	4.71723209
3	Raichur	5.1366686	5.25976625
7	Raichur	2.44207479	2.35750309
11	Raichur	1.94438139	1.79333694

Table VII: Koppal Yield Prediction

ID	Region	Actual Yield	Predicted Yield
4	Koppal	1.80415509	1.66117607
5	Koppal	2.44207479	2.35750309
6	Koppal	1.54003292	1.39066752
8	Koppal	2.0446274	1.95209759
9	Koppal	1.94438139	1.79333694

After we get the above dataset values for Raichur and Koppal, we calculate the average Yield of each region.

The Yield Prediction of Paddy for Raichur Region is:

$$\boldsymbol{Prediction\ (Raichur)} = \frac{\boldsymbol{Output}}{\boldsymbol{number\ of\ dataset\ rows}} = \boldsymbol{4.194\ tons}$$

Similarly, The Yield Prediction of Paddy for Koppal Region is:

$$\boldsymbol{Prediction\ (Koppal)} = \frac{\boldsymbol{Output}}{\boldsymbol{number\ of\ dataset\ rows}} = \boldsymbol{1.959\ tons}$$

B. Analysis using R

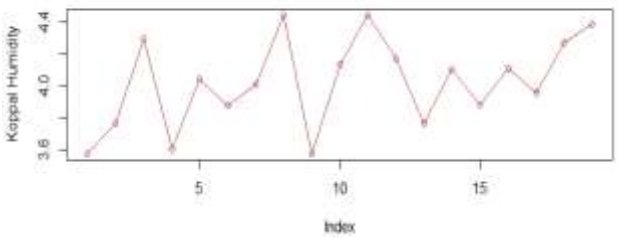

Fig.4: Humidity of Koppal District

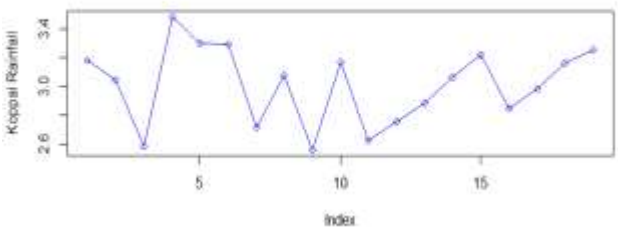

Fig.5: Rainfall of Koppal District

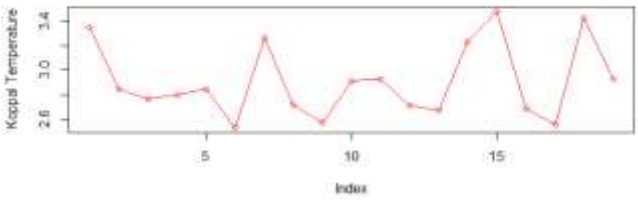

Fig.6: Temperature of Koppal District

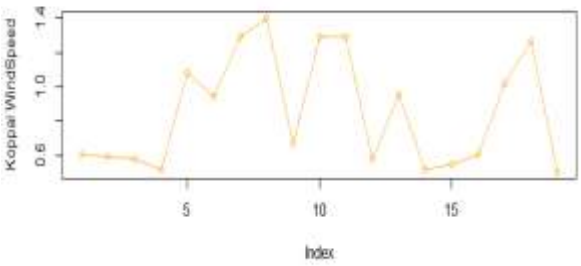

Fig.7: Wind speed of Koppal District

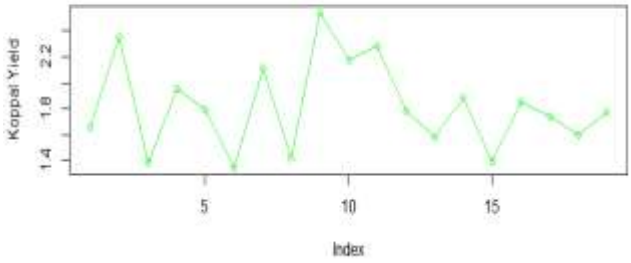

Fig. 8: Paddy yield of Koppal District

From the above graphs we can analyze that for both the districts when the humidity is high the yield is less. When the wind speed is very high the yield is less but more than the wind speed the major factor affecting the yield is temperature and rainfall along with humidity. Too much or too little rainfall and temperature has reduced the yield. The conclusion drawn from the analysis is that with low humidity, low or moderate wind speed, moderate rainfall and temperature the yield obtained is maximum.

C. The Performance of the System

The error rate helps in identifying the performance of the system by computing the percentage of error in prediction of yield.

Error Rate is given by:

$$[((\sum Predicted\ Yield) - (\sum Actual\ Yield)) \div (\sum Actual\ Yield)] \times 100$$

Error rate = 1.6%

IV. CONCLUSION AND FUTURE SCOPE

Conclusion

In this study, Apriori algorithm is used to estimate the productivity of paddy by the analysis of the factors effecting it from the input dataset. Here Hadoop clustering concept is used to analyze the relevant data such as rainfall, temperature, humidity and wind speed. By analyzing these data, the crop prediction is generated by which the end user such as farmers can make right decisions while farming. First map the complete input data and reduce using Map-Reduce framework.

After Mapping and Reducing is done, the data is classified in two districts Raichur and Koppal using KNN Classifier. Prediction is done using Apriori and for graphical analysis R is used.

V. FUTURE SCOPE

In the future work various other factors can be considered to improve the prediction and reduce the error rate. The scope of future work can be extended to:

- Consider other factors affecting yield like soil temperature, type of seed, soil pH etc.
- Reduce the error rate by using more efficient techniques.
- Using Spark for faster performance.

REFERENCES

1. Effect of Temperature and Rainfall on Paddy yield using Data Mining. Kuljit Kaur, KanwallPreet, Computer Science dept., Punjab University, Patiala, India. 506978-1-5090-3519-9/17/$31.00 ©2017 IEEE.
2. Soil Classification Using Data Mining Techniques: A Comparative Study by P. Bhargavi 1, Dr. S. Jyothi, MJR College of Engineering and Technology (Affiliated to JNTU, Anantapu), Piler. Andhra Pradesh, India, Department of Computer Science Sri Padmavathi Mahila Viswa Vidyalayam. (Womens University), Tirupati, Andhra Pradesh, India- IEEE 2016.
3. Big Data in Precision Agriculture: Weather Forecasting for Future Farming, M. R. Bendre, Department of Information Technology, SGGSIE &TNanded, India, Department of Computer Sci. and Engg. SGGSIE &TNanded, India Department of Instrumentation and Control. 978-1-4673-6809-4/15/$31.00 ©2015 IEEE.
4. Productivity Improvement in Agricultural Sector Using Big Data Tools. Ch Chandrashekar, Shekar, Dept of IT, AITAM, AP, India. 978-1-5090-6399-4/17 IEEE 2017.
5. Big Data Analytics Architecture for Agro Advisory System by Purnima Shah, Deepak Hiremath and Sanjay ChaudharySchool of Engineering and Applied Science,Ahmedabad University, Ahmedabad, India-IEEE 2016.
6. Towards Development of Spark Based Agricultural Information System including Geo-Spatial Data Purnima Shah School of Engineering and Applied Science, Ahmedabad, India purnima.shah@iet.ahduni.edu.in Deepak Hiremath School of Engineering and Applied Science, Ahmedabad, India deepak.hiremath@iet.ahduni.edu.in Sanjay Chaudhary School of Engineering and Applied Science, sanjay.chaudhary@ahduni.edu.in. 978-1-5386-2715-0/17/$31.00 ©2017 IEEE.
7. Research on Intelligent Acquisition of Smart Agricultural Big Data. Qiulan Wu, Yong Liang, Ying Li, School of Information Science and Engineering, China University of mining and Technology, China.
8. Research on Key Technologies of Intelligent Agriculture based on Agricultural Big Data, FengqiHao, Xuan Luo, Chunhua Mu, Shandong Provincial Key Laboratory of Computer Networks, Shandong Computer Science Centre (National Supercomputer Centre in Jinan), JiNan, ChinaMaize Institute of Shandong Academy of Agricultural Sciences, JiNan, China. 2016 International Conference on Smart City and Systems Engineering.
9. Big Data Analysis Technology Application in Agricultural Intelligence Decision System Ji-chun Zhao Beijing Academy of Agriculture and Forestry Sciences The Research Center of Beijing Engineering Technology for Rural Remote Information Services Beijing,China e-mail: zhaojichun_0@163com Jian-xinGuo Beijing Academy of Agriculture and Forestry Sciences The Research Center of Beijing Engineering Technology for Rural Remote Information Services Beijing,China 978-1-5386-4301-3/18/$31.00 ©2018 IEEE.
10. Survey on Weather Prediction using Big Data Analystics, P. Chandrashaker Reddy Computer Science and Engineering St. Martin's Engineering College Hyderabad, Telangana, INDIA, Dr. A. Suresh Babu Computer Science and Engineering JNTUA College of Engineering Ananthapuramu, AP,INDIA.Copyright.978-1-5090-3239-6/17/$31.00©2017IEEE.
11. Big Data in Weather forecasting: Applications and Challenges Ms. Himanshi Jain Department of Computer Science Jaypee University of Engineering and Technology Guan (M.P.), Indiadonica05@gmail.comMs. Raksha Jain, Department of Computer Science Gyan Ganga Institute of Technology and Science Jabalpur (M.P.), Indiarakshajain511@gmail.com _c 2017 IEEE.
12. Big Data analytics in agriculture and distribution channel Mukeshkumar MCA department SPTI College Andheri (W), Mumbai Prof.Mayuranagar MCA Department SPIT College Andheri (W), Mumbai 978-1-5090-4890-8/17/$31.00 ©2017 IEEE.
13. Exploring the Relationship between Maize Yield and Climate Big Data on Maize Belt of Northeast China Wang Dongjie, Li Zhemin, Wang Shengwei, Zhang Yongen, Xu Shiwei Agricultural Information Institute, Chinese Academy of Agricultural Sciences Key Laboratory of Agri-information Service Technology, Ministry of Agriculture, China Beijing, China. 978-1-5090-3619-6/17/$31.00 ©2017 IEEE.
14. Cloud Service Oriented Architecture (CSoA) for Agriculture through Internet of Things (IoT) and Big Data Pamidi Srinivasulu Department of Computer Science & Engineering DVR &Dr HS MIC College of Technology Kanchikacherla-521180, Krishna Dist., AP, India Venkat Department of Computer Science & Engineering &Dr HS MIC College of Technology Kanchikacherla-521180, Krishna Dist., AP, India (ICEICE2017).
15. IoT, Big Data Science & Analytics, Cloud Computing and Mobile App based Hybrid System for Smart Agriculture Sahitya. Roy, DrRajarshi. Ray, Aishwarya Roy, Subhajit Sinha, Gourab Mukherjee, Supratik Institute Of Engineering And management, Kolkata. 978-1-5386-2215-5/17/$31.00 ©2017 IEEE.
16. Big Data Approach in an ICT Agriculture Project, Dennis A. Ludena R. Department of Computer and Information Sciences Sojo University Kumamoto, JAPAN, AlirezaAhrary, Faculty of Computer and Information Sciences, Sojo University, Kumamoto, JAPAN, 2017 IEEE.
17. Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges by MOHSEN MARJANI1, FARIZA NASARUDDIN2, ABDULLAH GANI1, (Senior Member, IEEE), AND IBRAR YAQOOB1, Date of publication March 29, 2017.
18. Rice Yield Prediction Model Using Data Mining by Umid Kumar Dey, Computer Science and Engineering, East Delta University, Chittagong, Bangladesh. IEEE-2017.
19. X. Wu, X. Zhu, G. Wu and W. Ding, "Data mining with big data," inIEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 1, pp. 97-107, Jan. 2014.

20. S. Agarwal, "Data Mining: Data Mining Concepts and Techniques," 2013 International Conference on Machine Intelligence and Research Advancement, Katra, 2013, pp. 203-207.
21. Majumdar, Jharna, SnehaNaraseeyappa and ShilpaAnkalaki. "Analysis of agriculture data using data mining techniques: application of big data." Journal of Big Data4 (2017): 1-15.
22. S. Mishra, P. Paygude, S. Chaudhary and S. Idate, "Use of data mining in crop yield prediction," 2018 2nd International Conference on Inventive Systems and Control (ICISC), Coimbatore, 2018, pp. 796-802.
23. S. Sahni, "Ontology Based Agro Advisory System," Department of Computer Science and Engineering, IIT Mumbai, M.Tech thesis.
24. https://searchbusinessanalytics.techtarget.com/definition/big-data-analytics
25. www.esagu.in
26. http://www.bhoomi.karnataka.gov.in/
27. http://www.tcs.com/offerings/technologyproducts/mKRISHI/Pages/default.aspx
28. http://agropedia.iitk.ac.in/
29. http://cgwb.gov.in/District_Profile/karnataka/2012/KOPPA-2012L.pdf
30. https://data.gov.in/