International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 04 | Apr -2017 www.irjet.net p-ISSN: 2395-0072

Crop Plantation Prediction using Analysis of Weather Data

R. Aakash[1], N. Muthurasu[2] B.E., M.E., Karteek Rakshit[3], K. Ganapathy[4]

[1]Student, Dept. of Computer Science Engineering SRM University, Tamilnadu, India
[2]Assistant Professor, Dept. of Computer Science Engineering SRM University, Tamilnadu, India
[3]Student, Dept. of Computer Science Engineering SRM University, Tamilnadu, India
[4]Student, Dept. of Computer Science Engineering SRM University, Tamilnadu, India

---

Abstract - *Big data is the large collection of data that comes from numerous sources such as data from remote sensors, data posted on social media website, e-purchase transactions etc. Analyzing such data and using it for prediction analysis further on becomes tougher to process using older methods. The best method with which we can find some improvement is using Big Data Analytics (BDA) since these kind of data are again and again generated from many resources. Using BDA, we can find certain trends and patterns which will be useful in the upcoming stages of the project. In this project we are analyzing weather data and using the obtained result in predicting what kind of crops that can be grown under various weather conditions in all the districts in Tamilnadu using Hadoop ecosystems like HDFS, map reduce for analyzing and Fuzzy reasoning for prediction. The methods we use overcome the problems like limitation of data, data loss problem, processing one record at a time and further more limitations the previous methods had.*

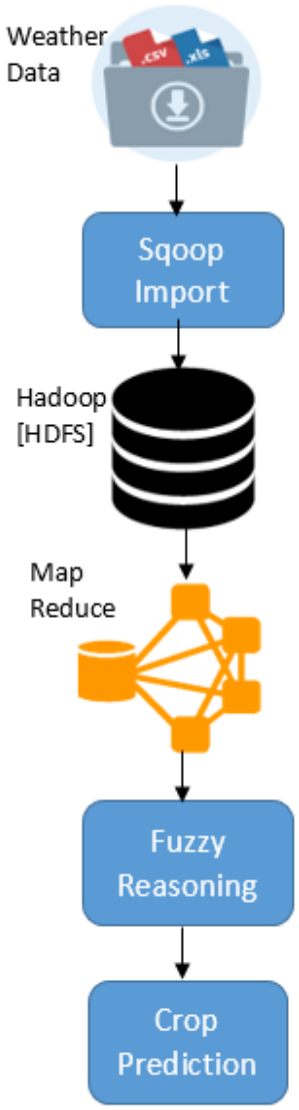
Key Words: **Big Data Analytics, HDFS, SQOOP, VM, Map Reduce**

1.INTRODUCTION

Agriculture is significant part of your country. Farmers use various manual methods to tell what crops can be grown at what time. With the advancement in the technology, the rainfall can be predicted based on various known methods. Various prediction methods are there as to what crops can be grown based on the weather conditions. The motive of this project is to integrate those methods and predict the crops that can be grown in certain conditions. This project utilizes Map reduce algorithm to analyze the weather data present from 2004 to 2015 in all the districts in Tamilnadu and using the obtained result the crops to be grown can be predicted using Fuzzy Reasoning which helps in yielding the maximum profit.

1.2 SYSTEM ARCHITECHTURE

This diagram explains various steps in the project



1.3 LIMITATIONS OF THE EXISTING SYSTEM

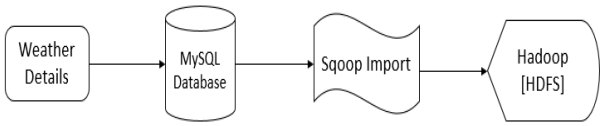
The previous system uses K-means clustering for both Analyzing and predicting the crops to be grown. Some of the limitations were, it takes more execution time, the inability to use other datasets to process like soil based data set and it can only run only one record at a time that is, taking into consideration only one year or one district at a time and maintenance cost was also high in the existing system. The proposed system on the other hand can run larger data sets the existing system could not.

2 METHODOLOGY

2.1 DATA PREPROCESSING AND MIGRATION USING SQOOP

The weather data containing information about the amount of rain in various districts of Tamilnadu collected from various sources cannot be directly accessed by the HDFS. In order to process the data further, the raw data has to converted into a suitable form supported by HDFS. Using MySQL Administrator, the weather data which is in the form of .csv (comma separated values) is taken backup and saved in a suitable place. This being the first part of the first module is over. Since we are processing the data using HDFS and Cloud-Era being an obvious choice for this since its open source and free has to be moved to the Cloud-Era Virtual Machine. The taken backup of the weather data is moved in to the VM and stored in the current working directory. This backup has to be moved into the HDFS for further operations, so using the inbuilt tool SQOOP we are moving the data into the HDFS. The reason why SQOOP is used is that since we are dealing with lots of data to analyze all of them at once is an impossible task so we are going to analyze it in parallel, for that the task would be so easy if the data had been stored in clusters in the HDFS. This is where SQOOP comes into place, using this tool automatically saves the data in clusters so that the analyzing part can be done in parallel, that is multiple VMs running at the same time and output is achieved faster than any of the traditional methods.

Figure 2.1.1 Importing using Sqoop



2.2 DATA ANALYSIS WITH MAP REDUCE

After successfully moving the data into the distributed file system, further operations can be started. Using Map Reduce algorithm the data can be analyzed and the required output can be obtained. Generally, Map Reduce is a processing technique based on java. The Map Reduce algorithm contains three important tasks, namely Driver, Mapper, Reducer. Driver is used to create jobs and configure Mapper and Reducer. Mapper takes a set of data and converts it into another set of data. Reducer takes the output from a map as an input. In this module of the project using the data obtained in HDFS, the minimum, maximum and average values of the rainfall for the years 2004-2015 is found for all the districts in Tamilnadu.

2.2.1 DRIVER

This is a part of Map Reduce where the jobs to be run by the mapper and reducer are created. The Mapper and Reducer is configured here. The input and the output parts are created. The Mapper and Reducer classes are created. The input format and output format of the Mapper and Reducer are specified here. This is in fact where the main () function resides and this is from where the mapper and reducer are called.

2.2.2 MAPPER

This part of the program is used to extract the data needed for further processing, since there is no use in sending all the data for every execution. The data to be sent is decided by the user. There are possibilities like sending the entire data so that we can get the overall rainfall in Tamilnadu for all these years or sending only the months the users wants for a specific district which is totally depending on the user

2.2.3 REDUCER

This is where the entire arithmetic operations take place. Considering this project, we are going to find the minimum, maximum and average values for all the districts in Tamilnadu. These operations are the key to find the required output which is further sent to the next module where the actual prediction takes place

Table 2.3.1: Sample Map Reduce Output

District	Minimum	Maximum	Average
Coimbatore	1.3	17.7	7.7285714
Cuddalore	0	73.3	20.833334
Dharmapuri	0	196.1	39.491665
Dindigul	0	100.5	17.125
Erode	0	218.7	47.416668
Kanchipuram	0	28.8	10.033334
Kanyakumari	0	114.2	20.625
Karur	0	43.8	6.9166665

3. CROP PLANTATION PREDICTION

This part of the project is done using python because of its easiness and robustness. After getting the output from map reduce, the output data is exported from HDFS to MySQL using SQOOP Export which works like the exact opposite of the SQOOP import command. From MySQL the data is saved in python which is used for further calculations. Since we are using Fuzzy Reasoning there should be classifications. High rainfall being one of the classifications and Low rainfall being the other. Using this and the data having crops specific to each districts which was already collected, fuzzy logic can be applied. Data having crops specific to each districts is present just to make sure that crops we get as output is possible to plant there. The output to be obtained depends on the user, the user has to select the district he wants and the month during which the crops has to be planted. This final part of the project will be able to return the average rainfall in that district, in that particular month and also the crops which can successfully be planted there which further helps us to yield the maximum profit

Figure 3.1.1 Average Rainfall

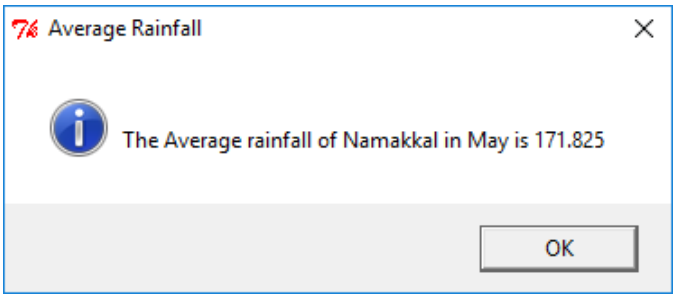
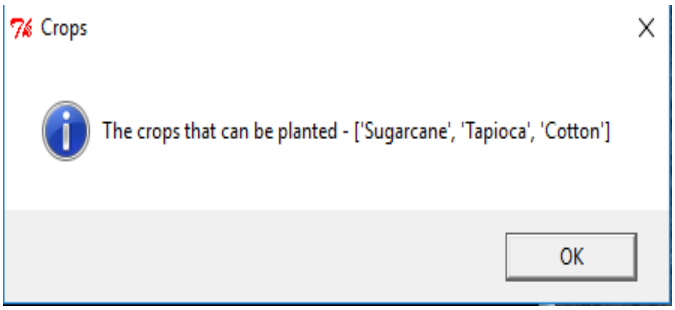


Figure 3.1.2 Crops that can be planted



4. FUTURE ENHANCEMENTS

Although the amount of data being analyzed and predicted is high, there is a scope of adding further more data. This project processes data about all the districts in Tamilnadu but in the future all the districts of our entire country can be processed, analyzed and predicted. Here only the rainfall data is used to predict the crop plantation but in the future even soil, humidity and further more commodities can also be added to make the project more accurate and useful foe people. Considering the methods used, in the future with the rapid advancements of technology day to day new methods can be incorporated to make this project more viable

5. CONCLUSIONS

With the increase in the advancements of technology, one should use it to their strengths. Year by year the complexity in prediction algorithms will increase, this is one step closer to attaining the perfect crop prediction algorithm which will be useful for farmers as well as fellow researchers who should help in further refining the program as well as increasing the robustness of the program.

REFERENCES

[1] N. Sundaravalli & Dr.A.Geetha "A Study & Survey on Rainfall Prediction And Production of Crops Using Data Mining Techniques"

[2] Yaser Jararweh, Izzat Alsmadi , Mahmoud Al-Ayyoub & Darrel Jenerette "The Analysis of Large-Scale Climate Data: Jordan Case Study"

[3] Awanit Kumar & Shiv Kumar "Prediction of Production of Crops using K-mean & Fuzzy Logic"