FINAL PROJECT (STAT 517)

Yadav, Rohit Kumar (yada6101@vandals.uidaho.edu)
Masters Computer Science, University of Idaho

Agricultural Analysis to solve the Drought conditions

Abstract

This Machine Learning Project is based on Agricultural datasets. Agricultural statistics and forecast is an important resource that the government has not explored commensurate to its impact. The aim of this project is to make this process computerized by implementing principles of data mining and analytics. More specifically, this project aims at targeting the social issue of drought, analysing data based on crop produce, amount of rainfall, agricultural inputs, irrigation, and similar factors for every crop in the state of Maharashtra, India.

Based on the research carried out in this project, effective countermeasures and suggestions will be given, which if implemented expeditiously, can help tackling the problem of drought in the state.

Data can be mined and analysed to find various trends and relations, such as – contrast between total irrigation area and type of crop; total principal and non-principal crop amount versus district-wise rainfall etc.

The end result of the project will be research based reports specifying these trends, studied and analysed from data taken over the past few years.

Introduction

In the current scenario, the government is collecting data only in its raw form, and this data is of no use to the end user, that is the farmers. Collecting this raw data, standardizing it, analysing it, and feeding it to a system that will provide relational trends is the aim of this project.

These relational trends will act as solutions for farmers, especially in drought afflicted areas. For example, the cultivation of Kharif Rice is resource intensive and should be only done in a rainfall rich period, otherwise it could use up the natural reserves of ground water, leading to deficiency in the water table, which consequently leads to drought like conditions. This example, as naïve as it may sound, represents a broad class of trends which when extrapolated efficiently using the existing data mining algorithms, can produce a richness of solutions, which when adhered to, will help alleviate the aforementioned drought like conditions.

The ultimate objective of this project is to stand as a system that when fed with data regarding various parameters, successfully produces trends and correlations that can help the user of the system develop solutions to tackle or minimize the damage of drought. Computerization of this process will drastically reduce the time required to study patterns and carry out extensive research to generate reports, and will give a close estimation of the required outcome.

FINAL PROJECT (STAT 517)

Yadav, Rohit Kumar (yada6101@vandals.uidaho.edu)
Masters Computer Science, University of Idaho

Problem Formulation and Solution

The practice of farming is one of the major occupations in our country, and a major produce of a variety of crops come from the state of Maharashtra. With an enormous agricultural sector in the state, the climatic conditions prevailing here are somewhat contradictory to what is required. Maharashtra is a drought ridden state, with one of the most devastating droughts occurring only in the recent past. Drought is a prolonged period of abnormally low rainfall leading to a scarcity of water for human use and farming. Data Mining is an emerging research field in agricultural crop yield analysis. In this project, our focus is on the applications of Data Mining techniques in agricultural field. Different Data Mining techniques are in use, such as K-Means, K-Nearest Neighbour (KNN) and Support Vector Machines (SVM) for very recent applications of Data Mining techniques in agricultural field.

The system will analyze data in two phase, using a classifier for each phase. Data will be used for both, training and testing purposes. The extracted data includes records for rainfall, temperature and pressure for eight districts of Maharashtra on an average-per-month basis. 80 percent of the data will be used for training the system, and the remaining 20 percent of the data will be used for testing the accuracy of the system.

I will use set of 4 datasets namely rainfall for rainfall in the different areas, temperature and pressures, crop for crop statistics and the governmental announced drought conditions in the areas.

For example Datasets Crops used in this project having 7 attributes namely District, Year, Season, Crop, Area, Production, Productivity having the total datasets of 514*7

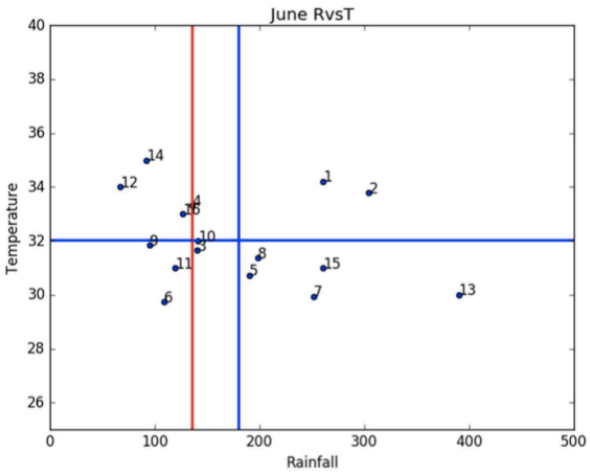
	District	Year	Season	Crop	Area	Production	Productivity
2	AMRAVATI	2001	Kharif	Bajra	1000	400	0.4
3	AMRAVATI	2001	Kharif	Jowar	109000	152800	1.40183486
4	AMRAVATI	2001	Kharif	Soyabean	153200	143600	0.93733682
5	AMRAVATI	2001	Rabi	Jowar	100	100	1
6	AMRAVATI	2001	Whole Year	Sugarcane	3000	204100	68.0333333
7	AMRAVATI	2002	Kharif	Bajra	800	500	0.625
8	AMRAVATI	2002	Kharif	Jowar	102900	144000	1.39941691
9	AMRAVATI	2002	Kharif	Soyabean	156500	150000	0.95846645
10	AMRAVATI	2002	Rabi	Jowar	100	100	1
11	AMRAVATI	2002	Whole Year	Sugarcane	2300	124000	53.9130435
12	AMRAVATI	2003	Kharif	Bajra	800	400	0.5
13	AMRAVATI	2003	Kharif	Jowar	100200	143100	1.42814371
14	AMRAVATI	2003	Kharif	Soyabean	155000	216100	1.39419355
15	AMRAVATI	2003	Rabi	Jowar	100	100	1
16	AMRAVATI	2003	Whole Year	Sugarcane	2500	155300	62.12
17	AMRAVATI	2004	Kharif	Bajra	700	400	0.57142857
18	AMRAVATI	2004	Kharif	Jowar	92800	88700	0.95581897
19	AMRAVATI	2004	Rabi	Soyabean	167100	80000	0.47875524
20	AMRAVATI	2004	Whole Year	Sugarcane	1400	58700	41.9285714
21	AMRAVATI	2005	Kharif	Bajra	300	100	0.33333333
22	AMRAVATI	2005	Kharif	Soyabean	181200	204400	1.12803532
23	AMRAVATI	2005	Kharif	Sugarcane	1100	72500	65.9090909
24	AMRAVATI	2006	Kharif	Bajra	200	100	0.5
25	AMRAVATI	2006	Kharif	Jowar	94700	111500	1.17740232
26	AMRAVATI	2006	Kharif	Soyabean	216700	323600	1.49330872
27	AMRAVATI	2007	Kharif	Bajra	2	1	0.5
28	AMRAVATI	2007	Kharif	Jowar	779	1058	1.35815148
29	AMRAVATI	2007	Kharif	Soyabean	2858	4857	1.69944017
30	AMRAVATI	2007	Whole Year	Sugarcane	12	959	79.9166667
31	AMRAVATI	2008	Kharif	Bajra	300	200	0.66666667
32	AMRAVATI	2008	Kharif	Jowar	40300	36700	0.91066998
33	AMRAVATI	2008	Kharif	Soyabean	373000	124200	0.33297587
34	AMRAVATI	2008	Kharif	Sugarcane	800	44500	55.625

Fig: Crop Datasets for crop statistics

FINAL PROJECT (STAT 517)

Yadav, Rohit Kumar (yada6101@vandals.uidaho.edu)
Masters Computer Science, University of Idaho

Premilinary Results:



Conclusions:

This project highlights the application of machine learning and data mining algorithms in the field of agriculture. Crop productivity and drought predictions, if presented in a proper format to the end-users, the farmers, it will immensely help drought afflicted villages and districts.