Prediction of Various Crops in Agricultural Field Using Decision Tree and Naviebayes Algorithm in Machine Learning

Mrs. B.Ganga Bhavani
Associate Professor
Department of
ComputerScience And
Engineering
B V C Engineering
College, Odalarevu

Mr. G.L.N.V.S.Kumar
Associate Professor
Department of MCA
Bonam Venkata
Chalamayya Institute Of
Technology and Science
Amalapuram, INDIA

Moram Lakshim
Rekha
Assistant Professor
Department of MCA
Bonam Venkata
Chalamayya Institute Of
Technology and Science
Amalapuram, INDIA

Mr. K N V P S
Brahma Ramesh
Assistant Professor
Department of
ComputerScience And
Engineering
B V C Engineering
College, Odalarevu

Abstract - **The Agriculture plays a dominant role in the growth of country's economy. Climate and other environmental changes has become a major threat in the agriculture field. Machine Learning (ML) is an essential approach for achieving practical and effective solution for this problem. For better crop yield the artificial Neural Network have demonstrated to be an effective tool for modeling and prediction by using Decision tree and Navi Bayes Algorithm.**

Keywords - CART, Decision Tree, Navie Bayes, Classification, Regression tree.

I. INTRODUCTION

Agriculture is truly a backbone of the Economy of India. Agriculture is truly a backbone of the Economy of India. Perhaps you may know that India is an agricultural dominated country, that's why most of the population here depends on agriculture. India is the second largest country in the world in terms of crop production. The importance of agriculture in rural areas of India is very much, which is one of the biggest and good means of living for their livelihood. India is today exporting very good crops to the world so India is a very good country when you are talking about the production of crops. Farming is done in different ways in different areas of India.

This paper mainly focused on the techniques and measures taken to improve farming by in calculating the technical knowledge and developments in order to make the agricultural sector more reliable and easier for the farmers by predicting the suitable crop by using Machine Learning techniques. By using the technology, precision agricultural technicians can break down a property in a very precise way so that what's planted on each inch of ground will have the optimal conditions for growth. This means looking at factors such as soil composition or water table levels. By using the existing conditions more efficiently, precision agricultural technicians. Across the globe India is the second largest country having people more than 1.3 Billion.

Many people are dependent on the agriculture and it is the main resource.

The data for the crop yield prediction is collected from Metrology Department of Kaggle. This is the annual data with all parameters of crops including states, crops, profit, yield. The aim of the proposed study is too effective and efficient in predicting which crop will be cultivated in which state with accuracy and precision.

Based on the crop yield and profit farmers can cultivate the suitable crops for particular states. This will increase the crop production and beneficial to the farmers and country 's economy.

Types of Crops:

Rice: Rice is the most important food crop of India covering about one-fourth of the total cropped area and providing food to about half of the Indian population. This is the staple food of the people living in the eastern and the southern parts of the country, particularly in the areas having over 150 cm annual rainfall. There are about 10,000 varieties of rice in the world out of which about 4,000 are grown in India

Wheat: Wheat is a grass widely cultivated for its seed, a cereal grain which is a worldwide staple food. The many species of wheat together make up the genus Triticum, the most widely grown is common wheat. Wheat is grown on more land area than any other food crop (220.4 million hectares, 2014).

Jowar : Sorghum popularly known as jowar is the most important food and fodder crop of dry land agriculture. The cereal crop is perennial in nature and possessing corn like leaves and bearing the grain in a compact cluster. Sorghum is the fifth most important cereal crop in the world after wheat, rice, maize and barley. It is found in the arid and semi aridparts of the world, due to its feature of being

Special Issue - 2021

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICRADL - 2021 Conference Proceedings

extremely drought tolerant. The nutritional value of sorghum is same as of that of corn and that is why it is gaining importance as livestock feed. Sorghum is also used for ethanol production, producing grain alcohol, starch production, production of adhesives and paper other than being used as food and feed.

Mustard : Mustard is the second most important and most prominent winter oilseed crop of India. It is grown mainly in the northern plains of India with some cultivated area in the eastern geography as well. It belongs to the group Cruciferae, with several cousin species cultivated. The others crops under the _Rapeseed & Mustard' category include Toria, Yellow Sarson, Brown Sarson, GobhiSarson or Canola and Black Mustard or Banarasi Rai. The small brown or yellow seeds contain up to 45 percent oil. The de-oiled cake is used as animal feed.

Sugarcane : Sugarcane is a tropical, perennial grass that forms lateral shoots at the base to produce multiple stems, typically 3 to 4 m (10 to 13 ft) high and about 5 cm (2 in) in diameter. The stems grow into cane stalk which, when mature, constitutes around 75% of the entire plant. A mature stalk is typically composed of 11–16% fiber, 12–16% soluble sugars, 2–3% non-sugars, and 63–73% water. A sugarcane crop is sensitive to climate, soil type, irrigation, fertilizers, insects, disease control, varieties, and the harvest period. The average yield of cane stalk is 60–70 tonnes per hectare (24–28 long ton/acre; 27– 31 short ton/acre) per year. However, this figure can vary between 30 and 180 tonnes per hectare depending on knowledge and crop management approach used in sugarcane cultivation. Sugarcane is a cash crop, but it is also used as livestock fodder

Types of Soils:

1. Alluvial soil: Alluvial soil is rich in nutrients and may contain heavy metals. These soils are formed when streams and rivers slow their velocity. The suspended soil particles are too heavy for the decreasing current to carry and are deposited on the riverbed. The finest particles are deposited at the mouth of the river, forming a delta. Alluvial soils vary in mineral content and specific soil characteristics depending on the region and geologic

2 . Black soil: This soil has high water holding capacity. So crops can be grown with less irrigation. It has high buffering and can hold nutrients in comparatively larger amount and for longer duration. This soil is very fertile and has been used traditionally for cotton cultivation leading to its name also.

Table 1: Different Crops in various states

0	state	Rice	Jowar	Bajra	Maize	Ragi	Wheat	Barley	Gram	Tur	Groundnut	Mustard	Soyabean	Sunflower	Cotton	Jute	Mesta	Sugarcane
1	Andhra Pradesh	2921.186	1054.115	906.734	3328.871	1202.798	792.5525	0	1139.672	441.4852	852.6383	356.35	1450.464	768.493	337.1242	0	1507.329	76472.37
2	Arunachal Pradesh	1190.479	0	0	1365.529	0	1641.025	0	1000	0	963.015	1258.428	0	0	0	0	0	19073.94
3	Assam	1448	0	0	710	0	1150	0	519	707	0	508	0	0	106	1736	909	37945
4	Bihar	1300.151	929.1806	933.166	2281.541	794.861	1876.421	1124.002	938.5558	1183.194	652.7778	853.5547	0	1365.018	0	1474.782	1344.15	43134.94
5	Chattisgarh	1177	843	0	1567	263	996	888	728	471	1122	369	857	459	136	0	361	2550
6	Goa	2652	0	0	0	1000	0	0	0	1813	0	0	0	0	0	0	0	52562
7	Gujrat	1610	901	1152	1300	868	2451	0	759	796	1161	1341	716	0	375	0	0	74050
8	Haryana	2894	296	1915	2238	0	3979	2735	725	988	809	1304	0	1598	452	0	0	59987
9	Himachal Pradesh	1447	0	0	2251	1104	1482	1287	901	0	0	495	1342	0	0	0	0	18013
10	Jammu & Kashmir	1960	589	571	1535	0	1543	631	0	0	0	635	0	0	0	0	0	0
11	Jharkhand	1413	988	1253	1465	632	1692	922	886	860	698	558	0	0	0	0	1051	34608
12	Karnataka	2561.393	945.2645	626.1794	2654.594	1492.443	736.6076	0	506.7264	489.3979	696.5809	278.5281	680.0272	456.3873	219.2412	0	265.7369	83235.97
13	Kerala	2197	480	0	1070	0	0	0	0	0	763	0	0	250	0	0	0	91363
14	Madhya Pradesh	862	985	1244	1525	951	1630	1228	855	754	992	925	928	453	164	0	382	58932
15	Maharashtra	1594	812	695	1835	992	1320	637	614	683	1066	317	1175	534	189	0	273	75204
16	Manipur	2315.246	0	0	2495.178	0	0	0	0	0	461.1111	0	0	0	0	0	0	32206.88
17	Meghalaya	1692	0	0	1452	0	1699	0	769	0	0	648	945	0	172	1430	835	0
18	Mizoram	1501	0	0	1814	0	0	0	0	0	742	1113	0	368	0	0	0	4076
19	Nagaland	1556	1246	1269	1609	0	1716	1791	1027	992	1308	842	1282	1197	570	587	0	44758
20	Orissa	1366.031	608.9923	559.4005	1412.742	641.3716	1445.9	0	627.3451	704.1701	1111.108	205.4544	769.2308	802.4356	527.1654	1797.091	796.46	60066.42
21	Punjab	3686	0	984	2702	0	4259	3309	892	876	868	1105	0	1602	563	0	0	60279
22	Rajasthan	1372	439	645	1314	0	2686	2263	652	606	1276	1043	1102	340	282	0	0	51264
23	Sikkim	1443	0	0	1488	895	1364	1216	0	0	0	696	827	0	0	0	0	0
24	Tamil Nadu	2889.749	816.9491	1217.121	2035.372	1685.402	0	0	671.9515	630.0695	1706.645	0	0	1258.19	281.7725	1440	0	101125.6
25	Tripura	2276	0	0	999	0	2326	0	744	978	789	0	0	0	224	1467	1370	47953
26	Uttarakhand	1943	0	0	1367	1398	1911	1090	926	833	1094	665	1073	0	0	0	0	58457
27	Uttar Pradesh	1964	965	1357	1427	1313	2634	2075	879	1054	741	1025	612	1530	188	1260	0	57343
28	West Bengal	2423.52	495.4724	0	2474.711	1137.009	2252.303	1129.416	871.2653	794.745	1595.666	875.3404	583.3333	1205.797	367.9162	2384.177	1639.423	73929.99
29	A & N Islands	2838	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	Chandigarh	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	D & N Haveli	1717	1200	0	0	0	1512	1867	0	813	0	0	0	0	0	0	0	0
32	Daman & Diu	1980.882	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	Delhi	2802	890	1756	0	0	3901	1000	0	0	0	700	0	0	0	0	0	0
34	Lakshadweep	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
35	Puducherry	2510	0	1000	0	0	0	0	0	0	1806	0	0	0	4167	0	0	85764

Soil	Code
Alluvial	AL
Red	RD
Black	BK
Mountain	MN
Laterite	LT
Desert	DT

Table 2: Different soils

States	Production	Profit
Andhra Pradesh	Rice	63851.8391
Andhra Pradesh	Jowar	10974.06918
Andhra Pradesh	Bajra	7414.478167
Andhra Pradesh	Maize	31369.83628
Andhra Pradesh	Ragi	5636.376098
Andhra Pradesh	Wheat	0
Andhra Pradesh	Barley	-1
Andhra Pradesh	Gram	5058.972382
Andhra Pradesh	Tur	0
Andhra Pradesh	Groundnut	10177.46897
Andhra Pradesh	Mustard	0
Andhra Pradesh	Soyabean	7632.152855
Andhra Pradesh	Sunflower	9739.718468
Andhra Pradesh	Cotton	0
Andhra Pradesh	Jute	-1
Andhra Pradesh	Mesta	1609.930873
Andhra Pradesh	Sugarcane	893130.4967
Arunachal Pradesh	Rice	999.586564
Arunachal	Jowar	-1

Pradesh		
Arunachal Pradesh	Bajra	-1
Arunachal Pradesh	Maize	4482.932344
Arunachal Pradesh	Ragi	-1

Table3: Production Vs Profit in states

Arunachal Pradesh	Rice	23809.59
Arunachal Pradesh	Maize	9558.702
Arunachal Pradesh	Wheat	20102.56
Arunachal Pradesh	Tur	20000
Arunachal Pradesh	Mustard	20223.32
Arunachal Pradesh	Soyabean	24841.36
Arunachal Pradesh	Sugarcane	111771.5
Assam	Rice	47911.5
Assam	Maize	4970
Assam	Wheat	14087.5
Assam	Gram	9866.19
Assam	Tur	14140
Assam	Mustard	17145
Assam	Cotton	2676.5
Assam	Jute	56854
Assam	Mesta	13635
Assam	Sugarcane	222357.7
Bihar	Rice	26003.02
Bihar	Jowar	7990.953

Table 4 : Crop Vs Profit

II. CART ALGORITHM

2.1 The Basis of CART Algorithm:

Classification and Regression Trees (CART) are one implementation of Decision Trees.The non-terminal nodes of Classification and Regression Trees are the root node and the internal node. The terminal nodes are the leaf nodes. Each non-terminal node represents a single input variable (x) and a splitting point on that variable; the leaf nodes represent the output variable (y). The model is used as follows to make predictions: walk the splits of the tree to arrive at a leaf node and output the value present at the leaf node.

The Classification algorithm is a Supervised Learning technique that is used to identify the category of new observations on the basis of training data. In Classification, a program learns from the given dataset or observations and then classifies new observation into a number of classes or groups. Such as, **Yes or No, 0 or 1, Spam or Not Spam, cat or dog,** etc. Classes can be called as targets/labels or categories.
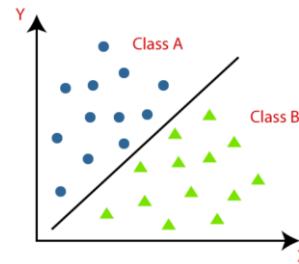
The algorithm which implements the classification on a dataset is known as a classifier. There are two types of Classifications:

Binary Classifier: If the classification problem has only two possible outcomes, then it is called as Binary Classifier.
Examples: YES or NO, MALE or FEMALE, SPAM or NOT SPAM, CAT or DOG, etc.

Multi-class Classifier: If a classification problem has more than two outcomes, then it is called as Multi-class Classifier.
Example: Classifications of types of crops, Classification of types of music.

III.NAVIE BAYES ALGORITHM

3.1 The Basis of naïve Bayes Algorithm:
Naïve Bayes algorithm is a supervised learning algorithm, which is based on **Bayes theorem** and used for solving classification problems.
It is mainly used in *text classification* that includes a high-dimensional training dataset.
Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.
It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.
Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles.
Bayes' Theorem:
Bayes' theorem is also known as **Bayes' Rule** or **Bayes' law**, which is used to determine the probability of a hypothesis with prior knowledge. It depends on the conditional probability.
The formula for Bayes' theorem is given as:

Special Issue - 2021

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICRADL - 2021 Conference Proceedings
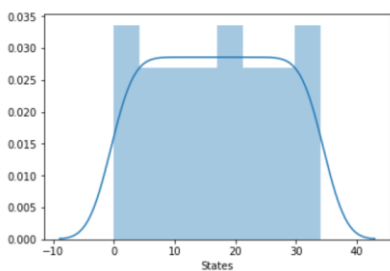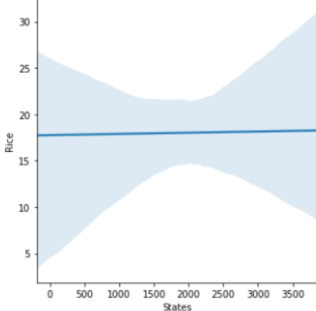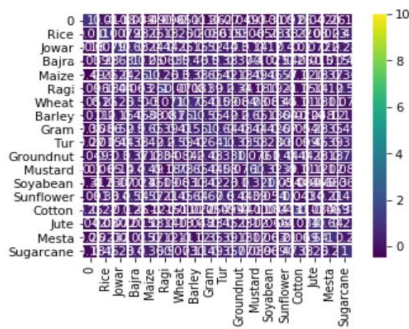
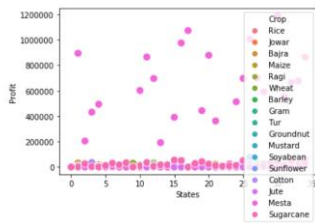$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where,

P(A|B) is Posterior probability: Probability of hypothesis A on the observed event B.

P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.

P(A) is Prior Probability: Probability of hypothesis before observing the evidence.

P(B) is Marginal Probability: Probability of Evidence.

Experimental Results:

Special Issue - 2021

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICRADL - 2021 Conference Proceedings
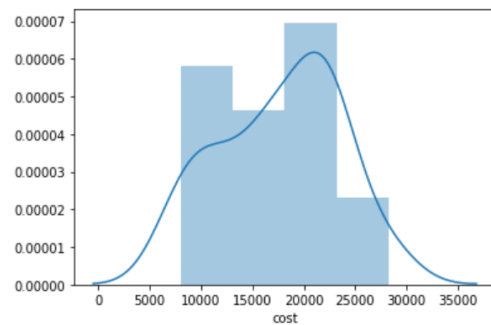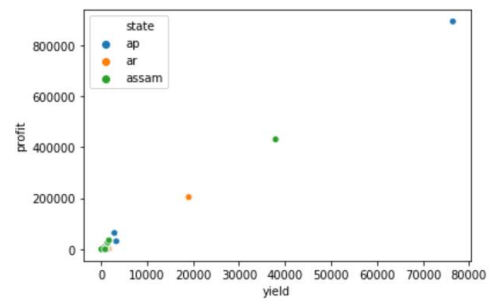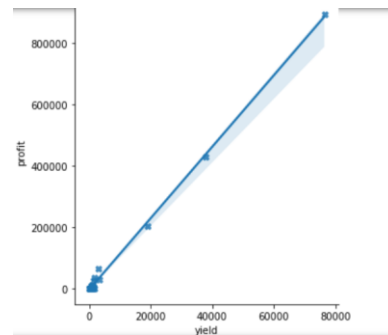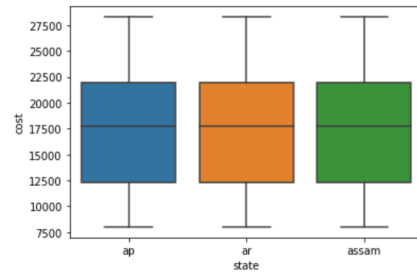
CONCLUSION

Based on the result of profit, soil test, yield we can generate specific hypothesis and by maintaining the root of the decision making we can easily recognize which crop is best in various states with the help of Decision tree Algorithm and Naïve Bayes Algorithm

REFERENCES

[1] N. Friedman, D. Geiger and M. Goldszmidt, "Bayesian Network Classifiers", *Machine Learning*, vol. 29, no. 2–3, pp. 131-163, 1997.

[2] Y. Ji and L. Shang, "RoughTree: A Classifier with Naïve-Bayes and Rough Sets Hybrid in Decision Tree Representation", *2007 IEEE International Conference on Granular Computing*, pp. 221-226, 2007.

[3] R. Abraham, J.B. Simha and S.S. Iyengar, "Medical Datamining with a New Algorithm for Feature Selection and Naïve Bayesian Classifier" in ICIT. 2007, IEEE Computer Society, pp. 44-49.

[4] J. Dougherty, R. Kohavi and M. Sahami, "Supervised and Unsupervised Discretization of Continuous Features" in Proc. 12th International Conference on Machine Learning. 1995, Morgan Kaufmann