



Article Title: Gradient Boosting and Naive Bayes Crop Yield Prediction and Fertilizer Recommendation

Gradient Boosting and Naive Bayes Crop Yield Prediction and Fertilizer Recommendation

Surya R¹, Sukanya ST²

¹Student, Department of MCA, Narayanaguru College of Engineering, Anna University

²Assistant Professor, Department of MCA, Narayanaguru College of Engineering, Anna University.

ABSTRACT

Farmers use Big Data to get information on changing Weather, Rainfall, Fertilizer Usage, Rainfall, and other factors that impact the crop yield. The yield of a crop is mainly determined by the climatic conditions like Temperature, Rainfall, Soil Conditions, and Fertilizers. All of this information assists farmers in making accurate and dependable decisions that maximize their productivity from cultivating the land. Recently, the Machine Learning Algorithms are used by the researchers to predict the yield of a crop before its actual cultivation. Firstly, Pre-process the data in a Python environment and then apply the Map Reduce Framework, which further analyses and processes the large volume of data. Secondly, K-means Clustering is employed on results gained from Map Reduce and provides a mean result on the data in terms of accuracy. Using Gradient Boosting Algorithm to predict the yield of crops based on the parameters like State, District, Area, Seasons, Rainfall, Temperature, and Area. To enhance the yield, this work study also suggests a fertilizer based on the soil conditions like NPK Values, Soil Type, Soil PH, Humidity, and Moisture. Fertilizer Recommendation is primarily done by using the Naive Bayes [NB] Algorithm.

Keywords: Agriculture, Big data analysis, Graphical visualization, k-means clustering, and map reduces framework, Gradient Boosting crop prediction, Naive Bayes fertilizer recommendation.

1 Introduction

Due to sudden changes in weather conditions, farmers and agriculture throughout the country suffer as they fail to produce enough crops. This leads them to take serious steps as they are unable to provide for their family and make ends meet. This also leads to a scarcity of availability of food resources in the country. The conditions of farmers in our country need to be changed. This paper aims at collecting and analyzing Temperature, Rainfall, Soil, Seed, Crop Production, Humidity and Wind Speed Data (in a few regions). India's economy is greatly influenced by agriculture as it serves as the backbone of the country. More than 50% of the country is dependent directly or indirectly on the agriculture sector and it is responsible for the employment of the major labour force of the country, which accounts for over 40%. Agriculture produces big volumes of data every year, and hence there is a need to get rid of the obsolete traditional predicting methods by charts and use the availability of the big data collected to



Article Title: Gradient Boosting and Naive Bayes Crop Yield Prediction and Fertilizer Recommendation

create a more prioritized and accurate predicting system. Big data will help confront the challenges and enhance the understanding of the whole sector. Big data analytics is the process of examining large data sets containing a variety of data types. The influence of weather can be deemed as a major priority in the prediction of crop yield. A lot of research work has been conducted in identifying how weather as a factor affects agriculture, but most of these studies require large complex information which is not directly available. This leads to the collection of data by estimation which can have either a negative or a positive effect. Hence improvement is needed in the methodology to compensate for the availability of data. This work focuses on crop prediction using various techniques.

Agricultural and meteorological data in India, which is mainly, collected from open dataset sources that contain information of crops from all states, but meteorological data revolves around three states and two union territories. As a result, the combination of all the data provides an elaborated view of the system and hence serves as the source of the big data. In this paper, a Map Reduce framework for data processing and a K-means clustering algorithm along with a recommendation function is carried out in the hope to propose crops to sow and elucidate big data applications in agricultural production. This paper arranged as follows recently, the Machine Learning Algorithms are used by the researchers to predict the yield of a crop before its actual cultivation. Farmers use Big Data to get information on changing Weather, Rainfall, Fertilizer usage, Rainfall, and other factors that impact the Crop Yield. The yield of a crop is mainly determined by the climatic conditions like Temperature, Rainfall, Soil Conditions, and Fertilizers. In proposed work, Firstly, Pre-process the data in a Python environment and then apply the Map Reduce Framework, which further analyses and processes the large volume of data. Secondly, K-means Clustering is employed on results gained from Map Reduce and provides a mean result on the data in terms of accuracy. Using Gradient Boosting Algorithm to predict the yield of crops based on the parameters like State, District, Area, Seasons, Rainfall, Temperature, and Area. To Enhance the Yield, this paper study also suggests a fertilizer based on the soil conditions like NPK values, soil type, soil PH, humidity, and moisture. Fertilizer Recommendation is primarily done by using the Naive Bayes [NB] Algorithm.

2 Recent Works

2.1 D. Bose, “Big Data Analytics in Agriculture

This paper talks about how Big Data Analytics combined with various structured and unstructured data helps in providing insight to farmers to make a decision as to which crops to grow and reduce losses due to unexpected or unpredictable disasters. In Section I the paper states that we can collect the data produced by sensors from the official databases that are usually maintained and governed by institutions. Here the author suggests we can collect and analyse the data in different stages in agriculture and see their influence in the big picture. It is



Article Title: Gradient Boosting and Naive Bayes Crop Yield Prediction and Fertilizer Recommendation

dependent on two major factors, the push and pull factor. Visualisation of agricultural data is done to simplify the complex, structured, and unstructured data. Interpretation of data can be done using methods like overviews, verifiable models, or in an Ad-Hoc manner and then visualized in the form of tables and graphs. In Section II, the paper talks about techniques like Predictive analysis where we can make the appropriate prediction of the future outcome on the basis of the previous data. A recommendation system is an informative system whose task is to offer an output that is based on functional patterns and behavioural data. Recommender systems generally give useful advice as the output is based on the approach used and the categories. The next method is Data Mining which can be defined as the process of extracting the previously unknown and useful information from large quantities of incomplete data for practical application. It plays a vital role in the agriculture sector, especially discovering patterns in big datasets, i.e., pattern mining next, the spike and slab regression analytic technique is discussed where the term spike and the term slab are used as a type of coefficient for regression. In the time series analytic technique using big data, time is taken as a variable that is independent with a motive to vegetation price movement, forecast crops and price fluctuation in the current market. In Section III, the implementation of analytic techniques in agriculture had been discussed. The first method is an intelligent crop recommendation system that considers all the factors such as soil conditions, temperature, rainfall and location. This system is further split into two different systems the crop predictor, whose main task is to help agriculturists by recommending crops and the rainfall prediction system that predicts the occurrence of rainfall for each month across the year. The next method discussed was Precision Agriculture using Map-Reduce used to allow variable rates and inputs which help in the understanding of time and space variability in criterion. Here the data is obtained and pre-processed. Then map-reduce is performed, and 3D visualization is done to visualize the output. Further crop prediction using various machine learning approaches was discussed. A few of them were 1) Grey wolf optimisation (GWO) technique 2) K-means clustering 3) Apriori algorithm 4) Naive Bayes. Next Smart Farming was discussed where a few of the services like Internet of Things, Cloud Computing, Mobile Computing were detailed about. The Crop analysis using Data mining techniques discussed is aimed at analysing greenhouse crops with the help of data mining techniques to extract patterns. With the help of the user interface and selection of specific greenhouse attributes, farmers will be able to predict yield patterns, crop patterns and further make important decisions based on them. Lastly, the author talks about a Spark-based system to perform collection, learning, training, validation and visualization of distributed data. This method of data analytics can be used for crop yield prediction, current weather trends and performing insights on Agricultural market data. In Section IV, the challenges that are faced in the analysis of big data in agriculture are discussed. The author states that obstacles faced for agriculture are usually Technical or Organizational problems. The paper further mentions the problems faced in the big data analysis of agriculture data,



Article Title: Gradient Boosting and Naive Bayes Crop Yield Prediction and Fertilizer Recommendation

majorly, availability, accessibility and scalability of data for analysis. Section V talks about the future scope of work where the author goes ahead and discusses various factors that could be helped with like product traceability, genetic engineering, supply chain, yield production, high precision, scientific simulations and so on and so forth. Lastly, Section VI contains a comparison table of big data techniques where one can notice that it suggests that we use Map Reduce for weather and climate data and K-Means Clustering for crop and vegetation data by collecting historical datasets.

2.2 B. R. Priya, D. Ramesh, E. Khosla, “Crop Prediction on the Region Belts of India: A Naïve Bayes map Reduce Precision Agricultural Model

In Section I of the paper, the focus is on the system of agriculture in Telangana. The data is collected from Crid as and farms of Hyderabad and Hayathnagar. A recommendation system recommends which crop to cultivate in the related seasons using Naïve Bayes classifier. Rice, Cotton, Maise and Chillies are the crops taken into consideration. Section II talks about the previous work done in the field of precision agriculture. The author tells the advantages and the grey areas of methods and models used in previous work like linear regression with neural networks, Map Reduce, KNN algorithm, a crop growth prediction model, sequential data assimilation. In Section III, the author describes the proposed methodology which is used to predict which of the four crops are suitable in Telangana. He talks about the modality and methodological conditions of 3 zones, i.e. (i) Northern Telangana, (ii) Central Telangana and (iii) in Southern Telangana, with seven major types of soil in which farmers mainly cultivate soybean, maise, rice, cotton where the water for irrigating the soil is provided by the rivers the Godavari and Krishna and monsoons (June-September). The suitable conditions for growing rice, maise and chillies are discussed .After collecting data from various sources like sensors from fields, images from satellites, data of crop, irrigation reports and weather data, it was pre-processed to find out the missing values and impute them using the mean method. Then feature selection and data extraction were performed in terms of soil, temperature, rainfall and atmospheric pressure. Further Map Reduce was implemented on this data, and then a Naïve Bayes classifier model for crop prediction was made using the Naive Bayes algorithm. This model recommended two or more crops based on the input data supplied. Section IV describes the results and recommends sowing and harvesting suitable crops.1) It was concluded that cotton should be planted in March/April as July to September are its ideal growth months where maximum growth is noted in the month of August. Since there is no noticeable growth from October to December, the crop can be harvested in January or February.2) Rice is grown in Rabi and Kharif season. It should be sown in July, as there is notable growth from August to September and it can be harvested in October and November. 3) Chilli requires good rainfall; hence the crop is sown at the start of July.4) The Maize plant should be sown at the end of June as it has the highest growth in July and is to be harvested in preferably September. Section V



Article Title: Gradient Boosting and Naive Bayes Crop Yield Prediction and Fertilizer Recommendation

discusses possible future enhancements. This work used Naïve Bayes to introduce a crop recommender system to make it very efficient when it comes to computation. The system can be used on a variety of crops as it is scalable.

2.3 C. Wu Fan, Chen Chong, Guo Xiaoling, Yu Hua, Wang Juyun, “Prediction of Crop Yield using Big Data

This paper discusses crop yield prediction, food security, Map Reduce and nearest neighbour modelling in terms of big data using agricultural data in China. In Section I, the paper talks about food security and its aspects like producing enough food and maintaining a stable supply of food in the market and how big data can help sort this out and points out that the earliest time in advance and accuracy are the priorities of predicting the crop yield. Section II portrays the advancement and application of big data in crop yield prediction. The paper states that effective plans for improving the performance of prediction of crop yield and the methods to take the maximum advantage of huge datasets related to agriculture and food security. Currently, big data can be obtained in semi-structural or non-structural forms from Recognition technology, Radio Frequency Identification, Remote Sensing, Weather stations. It is further reviewed by the paper that crop yield forecast is the most addressed topic, followed by climate change impact assessment and water resources. The well-developed methods have been categorised by Statistics methods, Remote Methods, Crop growth simulation, Econometrics. Section III proposes a model based on prior structure and weather data processing structure. First, the data was prepared by collecting it from the China Meteorological Administration with high accuracy of above 99%. Then Map Reduce was performed by partitioning the data into multiple sections. Then the map was executed according to certain rules followed by the Reduce function, where data having the same year was rearranged and combined. Output was written to distributed file systems. After that, weather similarity (defined by weather distances) was checked using nearest neighbours. The smaller was the distance between the two years which was quantified the similar the two years would be. at last, the autoregressive moving average model was used by combining two models. The output produced by one that has white noise as it is input which means that it has a linear relationship. In Section IV, the experiment is conducted on the already existing weather datasets, and the advantages of using this new method are talked through. This crop yield predictor is an application that has its basis on a processing structure that manages data in sequence to search for similar years. In the first step, the weather data was processed using Map Reduce, keeping precipitation, the intensity of sunshine and temperature at ground level as variables to calculate the daily mean and monthly mean. The process was divided into three steps, Map (to calculate monthly mean value), Reduce (to combine intermediate data), and storing the result. Next, a search for similar years was performed by conducting normalisation on three matrices to obtain a single 59×36 matrix, and then the difference of distance was obtained by computing the norm of the target year.



Article Title: Gradient Boosting and Naive Bayes Crop Yield Prediction and Fertilizer Recommendation

After sorting, 20 nearest neighbours were obtained similar to the target year. This was followed by preparing an ARMA model for prediction based on nearest neighbours found. This model was used to predict the crop yield of 2013 as an example and had a deviation of only 0.5%. The nearest neighbour's method using Map Reduce weather data processing structure had a balance of both accuracy and time in advance. Finally, in Section V, a conclusion is made that using the method mentioned above, an advantage of the already existing large datasets can be taken and put into use. Future possible work includes the faster accumulation of data and integrating weather calculation into the section that is processing data to reduce computing time. Lastly, this paper importantly focused on data mining in agricultural data from the perspective of time using a time aspect, the Map Reduce weather data processing structure. The same methods can be applied to different geographical aspects.

2.4 D. M. G. Ramya, C. Balaji, L. Girish, (2015) "Environment Change Prediction to Adapt climate-Smart Agriculture Using Big Data analytics"

The main aim of the paper is to predict changes in weather and help farmers in making agriculture-related decisions based on those changes. The paper has proposed a model to find solutions to modern world problems, such as worldwide food insecurity induced by frequent climate change, to predicting the impact of extreme weather events and mitigating its effect on global finance. They have made use of Big Data Analytics techniques to make an automatic prediction system. This paper builds the model based on the Hadoop framework. First, they collected data from various sources like social media, sensor data, weather forecasts etc. and loaded the pre-processed data into HDFS. HDFS stores datasets and provides backup features. They focused on three factors while collecting data, namely, precipitation, temperature and cloud cover for the state of Karnataka. The authors have mainly used Hive for reading and processing data. Hive's strong SQL skills make it possible to process huge volumes of data stored in HDFS. Hive converts SQL queries into a series of Map Reduce jobs. They have used Map Reduce to analyse the data and as an execution engine suitable for large data processing and to improve the response speed for returning query results. Then they implemented a prediction function for establishing forecast data through the k-means cluster algorithm. They used Apache Mahout to implement a logistic regression algorithm to predict the future based on the past data. For this, testing and training of data are performed. Then they evaluated the accuracy of the predicted result and represented the output using visualizations making use of the Flot end tool. Flot is a JavaScript plotting library. They used Pig script to perform analysis and the output was provided as an input to Flot end. They made various plots showing yearly and monthly average temperature for a particular region, maximum and minimum temperature, precipitation etc. The authors of the paper aim to improve their model such that it can be used for providing alerts in natural hazards in the future.



Article Title: Gradient Boosting and Naive Bayes Crop Yield Prediction and Fertilizer Recommendation

2.5 E. A. K. Kushwaha, S. Bhattacharya, “Crop Yield Prediction Using Agro Algorithm in Hadoop

This paper aims to predict the crop yield and suggest crops based on it which would, in turn, increase the profit of the farmers and overall, the entire agriculture sector. It also focuses on improving the quality of the crops using datasets for diseases. They have used a new algorithm called Agro Algorithm to predict the crop yield and suggest crops based on the crop yield and taking the soil type into consideration. In section I of the paper, the authors have used weather datasets containing information about temperature, rainfall in mm, wind speed, evaporation, humidity etc. They further used the weather datasets to determine the type of soil. They also used datasets for crop diseases to determine the ideal weather conditions which would be suitable for a particular crop to grow. Section II presents the numerous methods that already exist for crop prediction and their drawbacks. Here they discussed techniques like clustering, soft computing techniques such as k-means and artificial neural networks. In section III of the paper, the paper talks about some basic knowledge that is required to improve the quality of crops, such as selection of plant and soil factors such as pH, which would play an important role in getting a good yield. The properties of the soil should also be known beforehand. It is also important to select the right seeds and estimate the right amount of fertiliser and pesticides required. Section IV is the implementation. The implementation is performed on the Hadoop platform since the datasets are large. Normalisation is performed on the data stored in HDFS. This is done by taking the statistical average mean of data. First, the month in which the crop has to be sown is selected, then the classified data is used to predict the quality of soil and the recommended crop. The classification algorithm used is a simple statistical-based learning system. This prediction is represented using pie charts and this prediction is used to form five categories: “very good”, “good”, “average”, “bad” and “very bad”. Section V discusses the architecture used by the authors. They represented their architecture using a flowchart. Firstly, they collect multiple datasets related to agriculture and weather and perform the required analysis and classify the data. Then they used the classified data to predict the soil type and crop that can be sown. In section VI of the paper, the authors talk about some issues and obstacles that are faced in quality farming in India. These include technical gaps, small sizes of the farms, less availability of data and disparate harvesting systems. In section VII, they conclude that the crop yield is improved by using weather, soil, crop and disease datasets. This in turn, boosts the standard of production of crops. It helps farmers immensely in selecting a crop suitable for the weather and soil type. In the future, the authors aim to classify all the types of diseases for a particular crop and determine its cause which would further improve the quality of crops.



Article Title: Gradient Boosting and Naive Bayes Crop Yield Prediction and Fertilizer Recommendation

3 Proposed Work

Recently, the Machine Learning Algorithms are used by the researchers to predict the yield of a crop before its actual cultivation. Farmers use Big Data to get information on changing Weather, Rainfall, Fertilizer usage, Rainfall, and other factors that impact the Crop Yield. The yield of a crop is mainly determined by the climatic conditions like Temperature, Rainfall, Soil Conditions, and Fertilizers. In proposed work, Firstly, Pre-process the data in a Python environment and then apply the Map Reduce Framework, which further analyses and processes the large volume of data. Secondly, K-means Clustering is employed on results gained from Map Reduce and provides a mean result on the data in terms of accuracy. The soil conditions like NPK values, soil type, soil PH, humidity, and moisture .Fertilizer Recommendation is primarily done by using the Naive Bayes [NB] algorithm. Using Gradient Boosting Algorithm to predict the yield of crops based on the parameters like State, District, Area, Seasons, Rainfall, Temperature, and Area. To Enhance the Yield, this work study also suggests a fertilizer based on the soil conditions like NPK values, soil type, soil PH, humidity, and moisture.

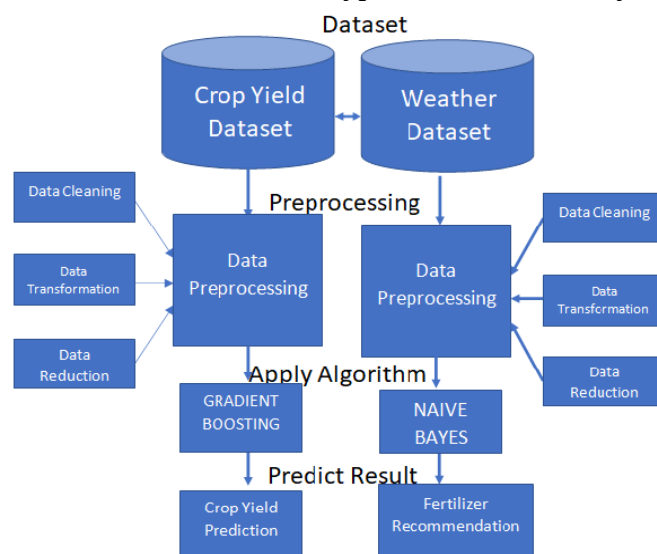


Figure 1: System Architecture

3.1 Load Dataset Module

The Prediction of Crop Yield is based on the Six Factor sand the recommendation of fertilizer is based on Six Factors. For Crop Yield Prediction State, District, Season, Crop, Area and Production. For Fertilizer Recommendation Temperature, Humidity, Moisture, Soil type Crop type and NPK.

3.2 Data Pre-Processing Module

- Getting the dataset
- Importing libraries
- Importing datasets



Article Title: Gradient Boosting and Naive Bayes Crop Yield Prediction and Fertilizer Recommendation

- Finding Missing Data
- Encoding Categorical Data
- Splitting dataset into training and test set
- Feature scaling

3.3 Gradient Boosting Crop prediction module

Gradient Boosting is a popular Boosting Algorithm in machine learning used for Classification and Regression Tasks. Boosting is one kind of Ensemble Learning Method which trains the model sequentially and each new model tries to correct the previous model. Gradient Boosting updates the weights by computing the Negative Gradient of the loss function with respect to the predicted output. Gradient Boosting Model is applied in Crop prediction system.

3.4 Naïve Bayes Fertilizer Recommendation Module

It is mainly used in Text Classification, which includes a High-dimensional Training Dataset. It is a Probabilistic Classifier, which means it predicts on the basis of the Probability of an Object. Naive Bayes' (NB) Model used in Fertilizer Recommendation System.

3.5 Data Analysis Module

Dataset is used to find the relation between produce per area and a particular crop using a Bar Graph and Scatter Plot for a Particular Region. The Elbow Graph has been plotted to find the number of clusters that should be made for a particular region's crops according to its produce per area. The Clustering Algorithm is then applied and the clusters are formed with the crops plotting their produce per area around the Cluster Centroids

3.6 Gradient Boosting Algorithm (Proposed Algorithm)

Gradient Boosting is also known as "Adaptive Boosting" algorithm which is suitable for both regression and classification problems. Gradient Boosting is one of the boosting techniques, itself an ensemble model and usually composed of a group of decision trees. In simple words, Gradient Boost combines weak learners into strong learners. Initially, the pre-processed original datasets are passed into a machine learning model Gradient Boosting, based on the dataset passed the prediction starts from the first iteration. In this dataset passed are separated into weak learners and strong learners. The weight of the weak learners is increased and combined with strong learners, then made into a newly updated dataset. This updated dataset is passed to the next iteration of the model. The process is done till the error gets minimum. This is known as boosting, the main aim is to improve the accuracy by minimizing the rate of error. The main reason to choose Gradient Boost is, it combines multiple classifiers to increase the accuracy. In other words, it combines multiple weak classifiers to build a strong classifier model so that the accuracy of the model gets increased. Gradient Boost is an ensemble model. An ensemble model makes better prediction and better performance compared to the activity



Article Title: Gradient Boosting and Naive Bayes Crop Yield Prediction and Fertilizer Recommendation

of a single model. It also works on fewer parameters and reduces the over fitting problem. The dataset containing information about Crop in India. Here use the gradient boosting algorithm to predict which crop to cultivate more production and profit for next year.

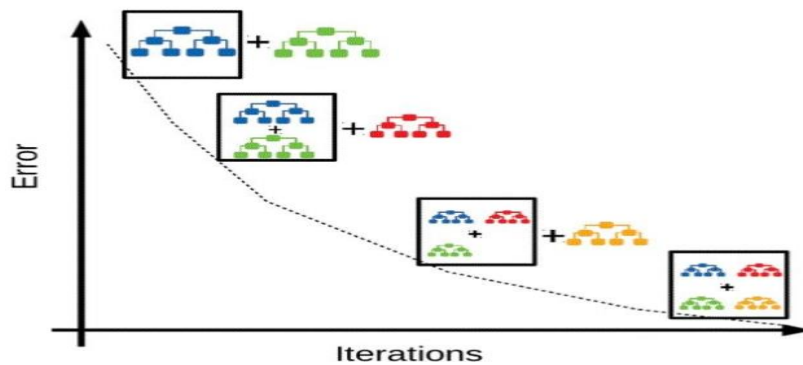


Figure 2: Prediction Graph

3.7 Gradient Boosting Steps

- Calculate the average of the target label
- Calculate the residuals
- Construct a decision tree
- Predict the target label using all of the trees within the ensemble
- Compute the new residuals

3.8 Naïve Bayes Algorithm (Proposed Algorithm)

Naive Bayes is a supervised machine learning algorithm that is used for both regression and classification problems. Naive Bayes itself is an ensemble method of problem solving. It combines a cluster of decision trees and the prediction process goes by, taking the average and voting of other models. The final prediction of the Regression problem is done by taking the average of other models and for the classification problem majority voting-based prediction process is under taken. It works on the concept of bagging. Initially, the original datasets are passed as a random sample to each decision tree with its placement called sampling. Hence each decision tree will produce its individual prediction. Based on the majority voting the output is finalized. Each subset of data is shared with each decision tree, so that the results produced have high accuracy. Comparing regression and classification problems, the algorithm works effectively in classification. The reason for choosing Naive Bayes is, also an ensemble model, it reduces over fitting problems and reduces variance which ends to increase the accuracy. It can handle missing values automatically and is very stable that is any change in one decision tree will not affect another one. Also, have less impact over the noise.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$



Article Title: Gradient Boosting and Naive Bayes Crop Yield Prediction and Fertilizer Recommendation

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.

3.9 Naive Bayes Steps

- Data Pre-processing step
- Fitting Naive Bayes to the Training set
- Predicting the test result
- Test accuracy of the result (Creation of Confusion matrix)
- Visualizing the test set result

4 Result and Discussion

- Here is used two datasets are used in this project. crop_production.csv and Weather.csv.
- This paper used two Machine Learning algorithms to predict the accuracy level of the algorithm. GB (Gradient Boosting)-Crop Prediction Analysis Naive Bayes [NB] Algorithm Fertilizer Recommendation.



Figure 3: Seaborn plot

5 Conclusion

The proposed system suggested a machine learning algorithm for crop yield prediction and for fertilizer recommendation. The proposed machine learning model successfully produces the output and both the algorithm Gradient Boosting, Naïve Bayes performs well. Gradient Boosting is more accurate compare to other models. Data training also faster especially on larger datasets, Naïve Bayes was fast and can be used to make real-time predictions.

Reference

1. Charvat Karel; Karel Charvat Junior; Tomas Reznik; Vojtech Lukas; Karel Jedlicka; Raul Palma; Raitis Berzins, Year: 2018, "Advanced visualization of big data for agriculture as part of



Article Title: Gradient Boosting and Naive Bayes Crop Yield Prediction and Fertilizer Recommendation

- databiodevelopment”, in Process. IEEE International Geoscience Remote Sensor Symposium, Vol: 67, no: 11, pp. 415–418.
2. Shah Purnima; Deepak Hiremath; Sanjay Chaudhary, Year: 2017, “towards development of spark based agricultural information system including geo-spatial data”, in Process. IEEE International Conference, Vol: 38, no: 12, pp. 3476–3481.
 3. Savla Anshal; Nivedita Israni; Parul Dhawan; Alisha Mandholia; Himtanaya Bhadada; Sanya Bhardwaj, Year: 2015, “Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture”, in Process International Conference Innovative Information Embedded Communication System, Vol: 5, no: 8, pp. 1–7.
 4. Ramya M. G. P. G; Chetan Balaji; L. Girish, Year: 2015, “Environment change prediction to adapt climate-smart agriculture using big data”, International Journal Computer Engineering Technology, Vol: 4, no: 5, pp. 1995–2000.
 5. Reddy P; Chandrashaker A; Suresh Babu, Year: 2017, “Survey on weather prediction using big data analytics”, in Process 2nd IEEE International Conference Electronic Computer Communication Technology, Vol: 9, no: 6, pp. 1–6.
 6. Vijayabaskar P. S; R. Sreemathi; E. Keertanaa, Year: 2017, “Crop prediction using predictive analytics”, in Process International Conference Computer Power Energy Informative Communication, (ICCPEIC), Vol: 40, no: 15, pp. 370–373.
 7. Rajeswari S. K. R. K. A; Kannan Suthendran; K. Rajakumar, Year: 2017, “A smart agricultural model by integrating IoT mobile and cloud-based big data analytics”, in Process International Conference Intelligent Computer Control, Vol: 32, no: 10, pp. 1–5.
 8. Ramya M. G. P. G; Chetan Balaji; L. Girish, Year: 2015, “Environment change prediction to adapt climate-smart agriculture using big data”, International Journal Computer English Technology, Vol: 4, no: 5, pp. 1995–2000.
 9. Kushwaha Ashwani Kumar; Sweta Bhattacharya, Year: 2015, “Crop yield prediction using agro algorithm in Hadoop”, International Journal Computer Science Information Technology, Vol: 5, no: 2, pp. 271–274.
 10. Dutta Pinky Saikia; Hitesh Tahbilder, Year: 2014, “Prediction of rainfall using data mining technique over Assam”, Indian Journal Computer Science Engineering, Vol: 5, no: 2, pp. 85-90.