CONSTRICTING THE BIVARIATE ANALYSISOFHIGHEST CROP YIELD PRODUCTION BASED ON DIFFERENT ZONES OF INDIA

Sasmita Kumari Nayak

Computer Science and Engineering, Centurion University of Technology and Management, Odisha, India
Email: nayaksasmita484@gmail.com

Abstract

In analysis of crop yield production, emerging research fields are Machine Learning Model and Data Mining. In agriculture, production of crops is a very much complex issue. Also, it is a big issue for farmers. Analysis of crop yield production is a highly essential step for predicting the production of crop. This is an initial step of this issue as well as for all types of problems. The analysis step is used to analyse for cultivation of which crops and what actions should be taken whilegrowing season of the yields, which helps to the farmer for production of crop. In this article, the analysis has been done for crops in terms of different zones, state wise top highest crops and their comparisons. The outcomes of simulation illustrate the prediction and cultivation of crops that helps to farmer for cultivating which crop in which region of India, so that the farmers will be benefited.

Key words: Bivariate Analysis, Crop Yield Analysis, Data pre-processing, Exploratory Data Analysis (EDA).

Introduction

Data mining and machine learning model are an analytical tool used by users to analyse data and to show the relationships among them.Farmers as well as Growers have got benefitted from prediction of yieldsfor making financial and managementdecisions (Khaki et al., 2019; Horie et al., 1992). Hence, in global food production, the much attention has given to Crop yield prediction (Khaki et al., 2019). Still, crop yield prediction is highly complexbecause ofvaried complex factors like temperature, weather condition, soil conditionand so on.Each crop has various attributes or parameters to get predictions with the help of various models and these models could be examined by doing many studies (Medar et al., 2019). Several Machine Learning (ML) models shall apply for getting the maximal production of crop. But, before applying Machine learning as well as data mining tool, analysis of crop data is highly required, which is the main purpose of this study. After analysing this, the specific problem will reach to some point to solve easily. Hence, the crop production will also reach to some point to predict the crop, which is dependent on weather conditions (cloud, temperature, rainfall and humidity) as well as geographicalconditions (depth areas, hill areas, river ground)(Medar et al., 2019).

From the various studies of agriculture field, I got multiple ways of increasing the economic growth of India. In analysis of crop yield production, an emerging research field is used named as Data Mining. The issue of prediction of crop has a great importancein agriculture field. All farmers have only one interest i.e. the expectation of high quantity of crop production. Earlier, the crop prediction can be done depending on the experience of farmers with specific crop and field. And this is a major issue, which can be solvable with the help of available data and Data mining techniques. Basically, data mining is a method for analysing data from varied viewpoints and summarized the same into important information.

This study analysed and implemented for predictionof crop by using the previous data. The outcomes of simulation illustrates that which zone of India and what cropswill be cultivated. So that farmers will be benefited much.

The organization of the paper is as follows: the related work of crop yield production and proposed methodology is presented in section II and III respectively. Section IV deals with the experimental outcomes followed by the conclusion and future work in section V.

Literature Survey

A huge number of crop yield prediction works are completed by different researchers with different models and found their accuracies by comparing the several ML methods in the prediction of leaf disease. A few of them are outlined in Table I.

Table I: Summary of crop yield prediction by using ML models.

Authors	Applications	Algorithms	Remarks
SitiKhairunniza-Bejo, SamihahMustaffha, Wan Ishak Wan Ismail (Medar et al., 2019; Siti et al., 2014)	Providingresults to some problems of farmers forfinding good yield	ANN	Time consuming process
AnshalSavla, HimtanayaBhadada, Vatsa Joshi, ParulDhawan (Medar et al., 2019; Anshal et al., 2015)	Based on parameters, understandand analyzecrop yield rate for zones	Classification, Clustering, Normalization	only provides framework
Raorane A.A, Dr.Kulkarni R.V(Medar et al., 2019; Raorane et al., 2015)	Rain fall estimation and reason investigation to get lower yield	Regression analysis	Not specified any specific method
B Vishnu Vardhan, D Ramesh	Analyze and verify the existing data based on multiple linear regression method	Multiple Linear Regressions	Less accuracy
Subhadra Mishra, Debahuti Mishra, GourHariSantra (Medar et al., 2019; Gour et al., 2016)	Forecastand increase crop yield rate	ANN, Regression analysis Decision Tree	Not specified any clear method
Karan deep Kauri (Medar et al., 2019; Karan et al., 2016)	Increasing the farming sector in the countries	ANN, Bayesian Belief Network, Decision Tree, Clustering, Regression analysis.	Less accuracy
Nishit Jain, Amit Kumar, SahilGarud, Vishal Pradhan, PrajaktaKulkarni (Medar et al., 2019; Nishit et al., 2017)	Predictcrop sequences and maximize yield rate and make benefits to the farmers. Also predictcrop diseases, study crop simulations, different irrigation patterns.	ANN, SVM.	Exact accuracy is not specified.
Elavarasan et al. (Klompenburg et al., 2020; Elavarasan et al., 2018).	A survey of crop yield prediction based on climatic parameters.	ML models	looking broad to get more attributes that represent crop yield
Liakos et al., (Klompenburg et al., 2020; Liakos et al., 2018).	A review on applications of ML in the agriculture.	ML models	Analysis on soil, water, livestock and crop management
Li, Lecourt, Bishop (Klompenburg et al., 2020; Li et al., 2018).	a review to determine the ripeness of fruits	ML models	decide the optimal yield prediction and harvest time
Beulah (Klompenburg et al.,	a survey on prediction of	Data mining	solved crop yield

2020; Beulah, 2019).	crop yield	techniques	prediction based on data mining techniques

Proposed Methodology

Generally, data mining and machine learning model are an analytical tool used by users to analyse data and to show the relationships among them. This research has been done using several Machine Learning algorithms, namely, SVM(Sasmita et al., 2020;SasmitaKumaraiNayak et al., 2020; Sripada et al., 2020), Decision Tree(Sasmita et al., 2020; Sripada et al., 2020), and Random Forest (Sasmita et al., 2020; Tapas et al., 2020; Sripada et al., 2020). ML techniques have been traditionally applied to large, highly dimensional databases. Machine learning (ML) is a subset of computer science, whereby a computer algorithm learns from prior experience. The steps of machine learning model as shown in Figure 1(Sasmita et al., 2020). The most essential part of machine learning model is the collection and pre-processing of data. This model has been applied to clean, normalize and pre-process the collected data called as crop yield dataset.

It is a time consuming process but also an essential for the model is need to understand the process of collecting, storing, transforming, reporting the data.

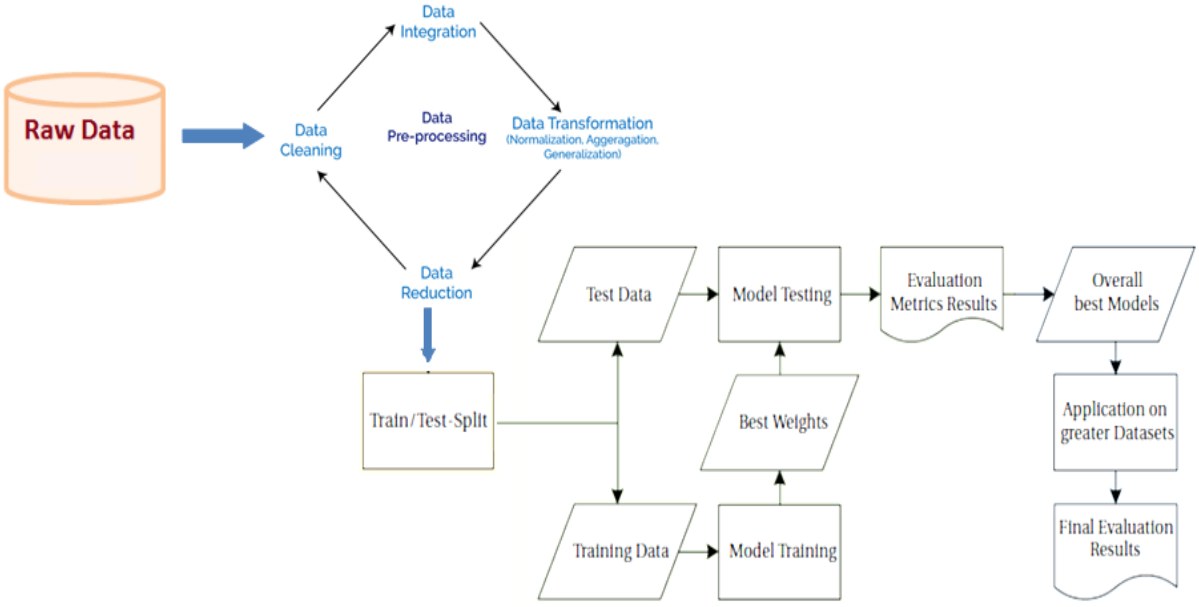

Figure 1: Steps of Machine Learning Model (Sasmita et al., 2020).

A. *Dataset Collection*

In this article, we have collected the crop production in India data from the website https://www.kaggle.com. It has around 17 years (from 1997 to 2014) of different crop yield productions from different states and districts. In total, the raw data consists of 7 columns or attributes and 246091 numbers of instances. Table II shows the description of crop yield data.

Table II: Description of Crop Production in India Data

Attribute Name	Type of Attribute	Description of Attributes
State_Name	String	Name of the State
District_Name	String	Name of the District
Crop_Year	Numerical	Year of the crop production
Season	String	Current season of the crop production

Crop	String	Type of crop
Area	Numerical	Area of the agricultural field
Production	Numerical	Total crop production

B. *Data Pre-processing*

In Data Science, data pre-processing is anessential step. It is a first step of machine learning model. In this case, data will clean, transform and normalisedform. To access variedattributes, work on missing data (either it will be deleted or imputed with suitablevalues) and verify the accuracy of data collected are the steps of data pre-processing. Outliers are data points which will be flagged as well as investigated once they are helpful for including in the analysis. This step cleans the given data so that it could be applied in the problem without any hassle.Data pre-processing is required to get the better accuracy for crop prediction. In the data cleaning phase, remove the missing vales from raw data. In this study, no transformation phase is required.After pre-processing of data, the number of instances will be 242361.

C. *Exploratory Data Analysis (EDA)*

Exploratory data analysis (EDA) is a way to analyseand summarise the collected data,which helps toget essential attributes or parameters for visualization purpose. In EDA, identification of variable is the initial step to build a predictive model.This can be performed by using the three methods, such as, univariate, bivariate and multivariate analysis. This study is focused on bivariate analysis. Bivariate analysis is implemented to get the association between each feature in the dataset and the target variable.

Result

This experiment is conducted on crop yield production data of India from 1997 to 2014. In this dataset, contains 246091 numbers of instances or records with 7 attributes of each. After data pre-processing, I have chosen only 242361instances for prediction. Hence, we have applied all the steps of data pre-processing through the implementations of ML models.

From the existing dataset initially created different zones like East, West, North, South, Central zone and so on with the categories of crops. From this information of zone, selected four zones, they are, East, West, North andSouth zone of India. From the analysis, I have got the various visualizations, which have discussed below.

Figure 2 represents the total production of crop according to the zone wise. This visualization provides the top zone of India to cultivate the crops. It clearly says that South zone of India cultivate more crops as compared to the other zones. Next discusses the topmost crop productions of each zone with state wise.

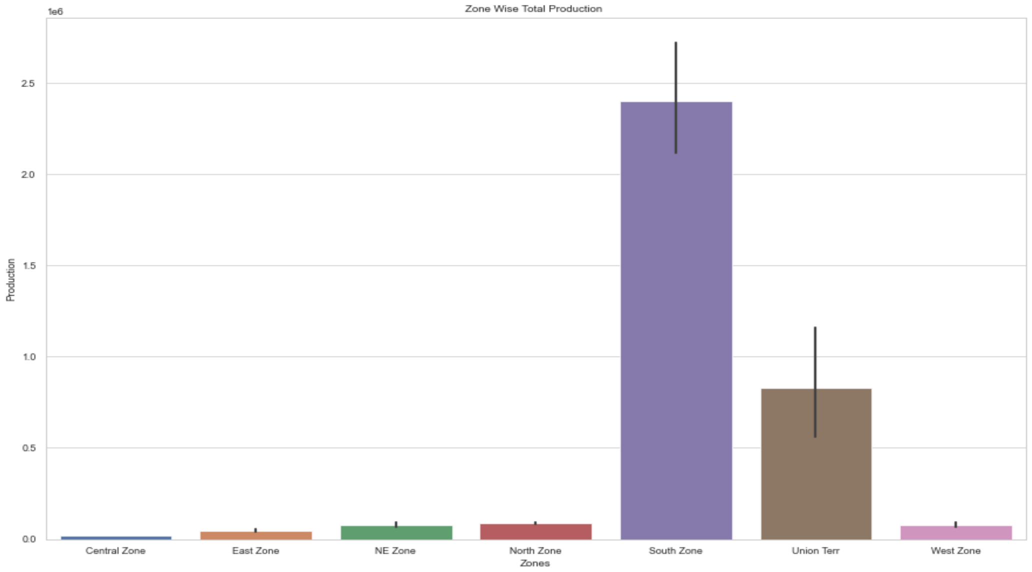

Figure 2: Total Crop Production as per Zone of India

In South zone of India, the top three most states for crop productions are Kerala, Andhra Pradesh and Tamil Nadu as shown in figure 3.

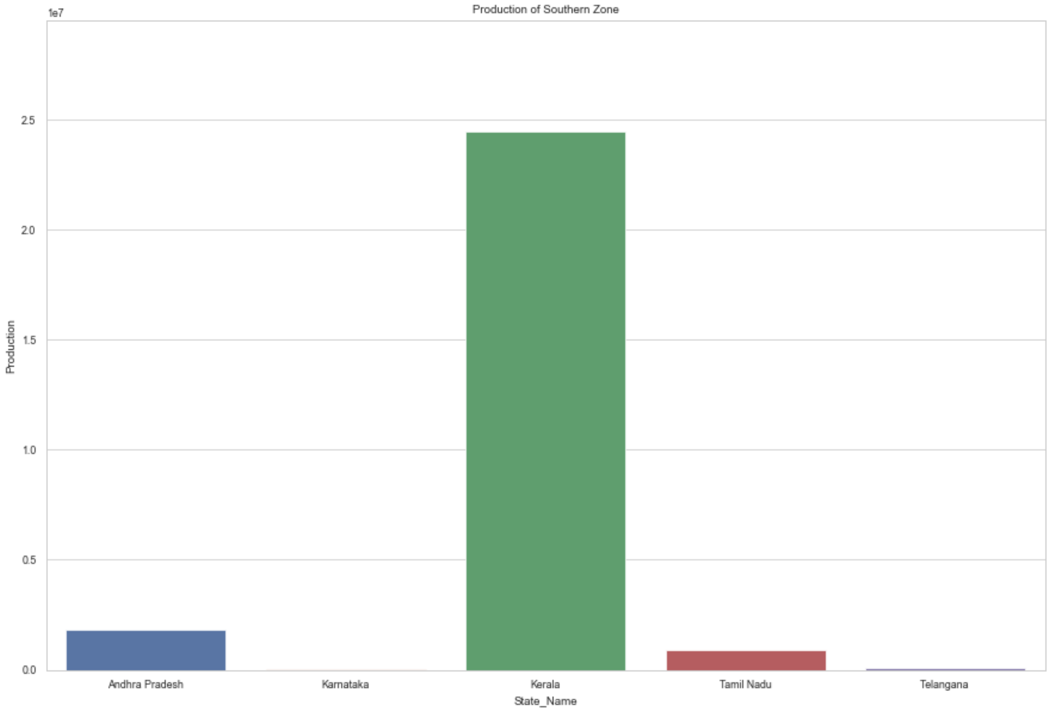

Figure 3: State wise crop production of South Zone of India

Similarly, for East, West and North zone are illustrated in figure 4, 5 and 6 respectively. Table III describes the top 3 most states for each zone of India.

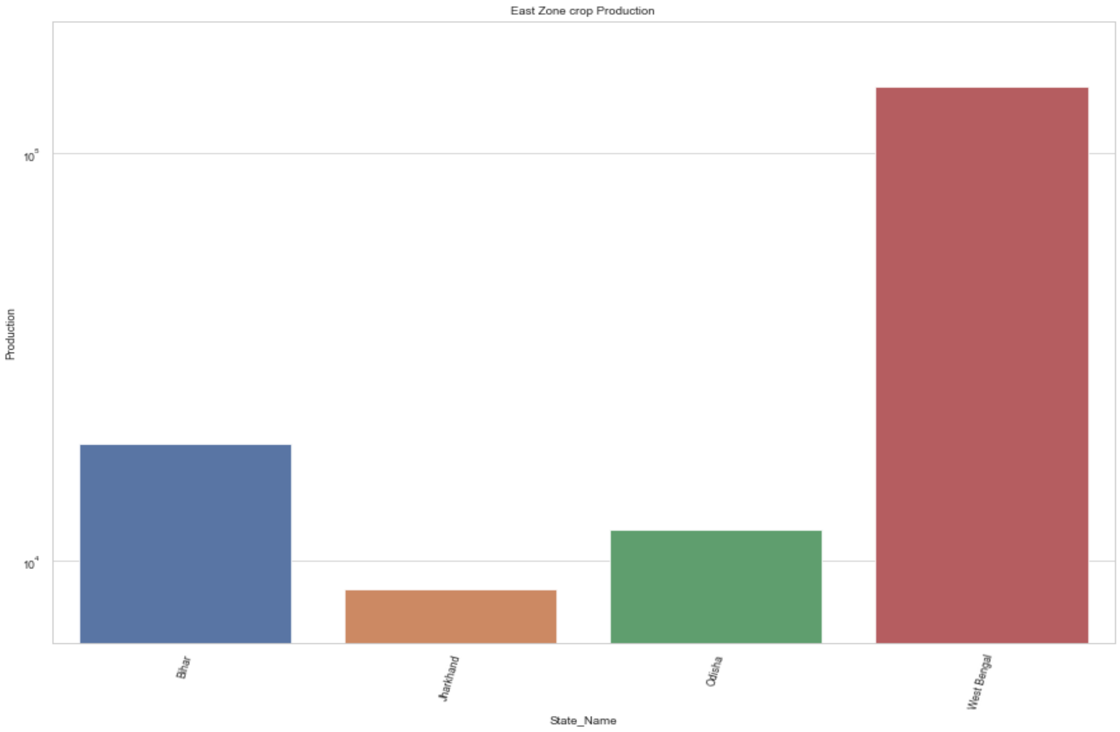

Figure 4: State wise crop production of East Zone of India

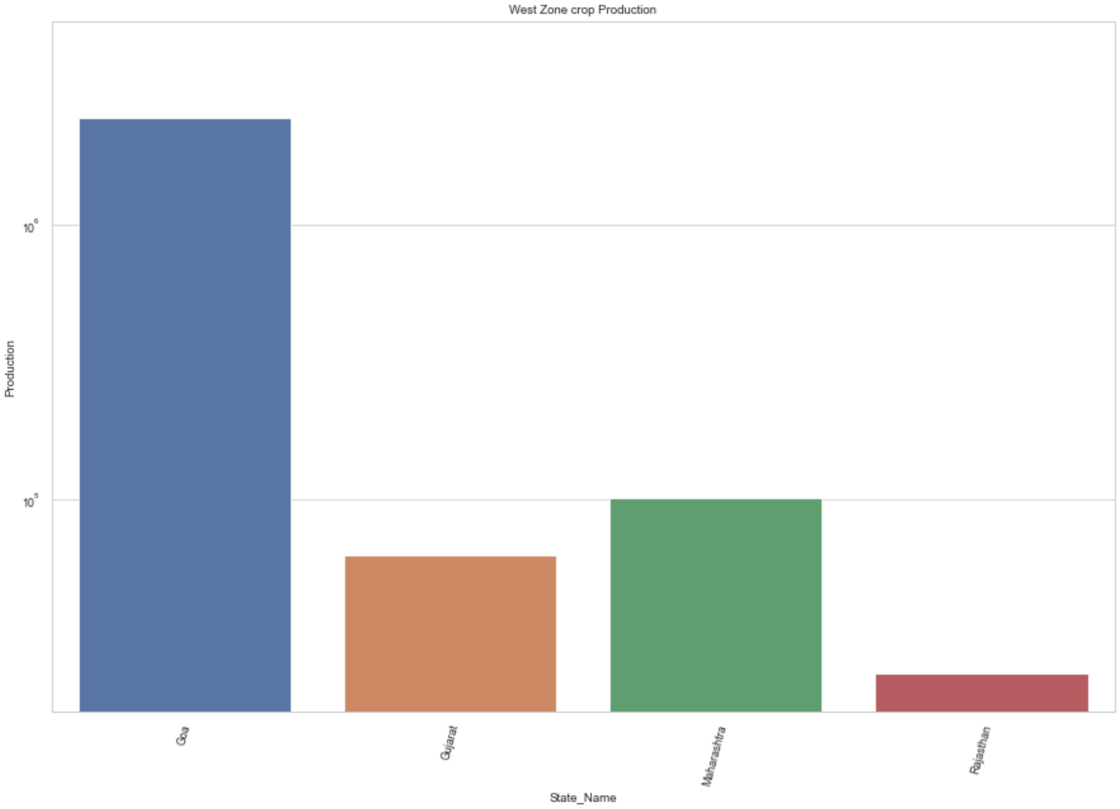

Figure 5: State wise crop production of West Zone of India

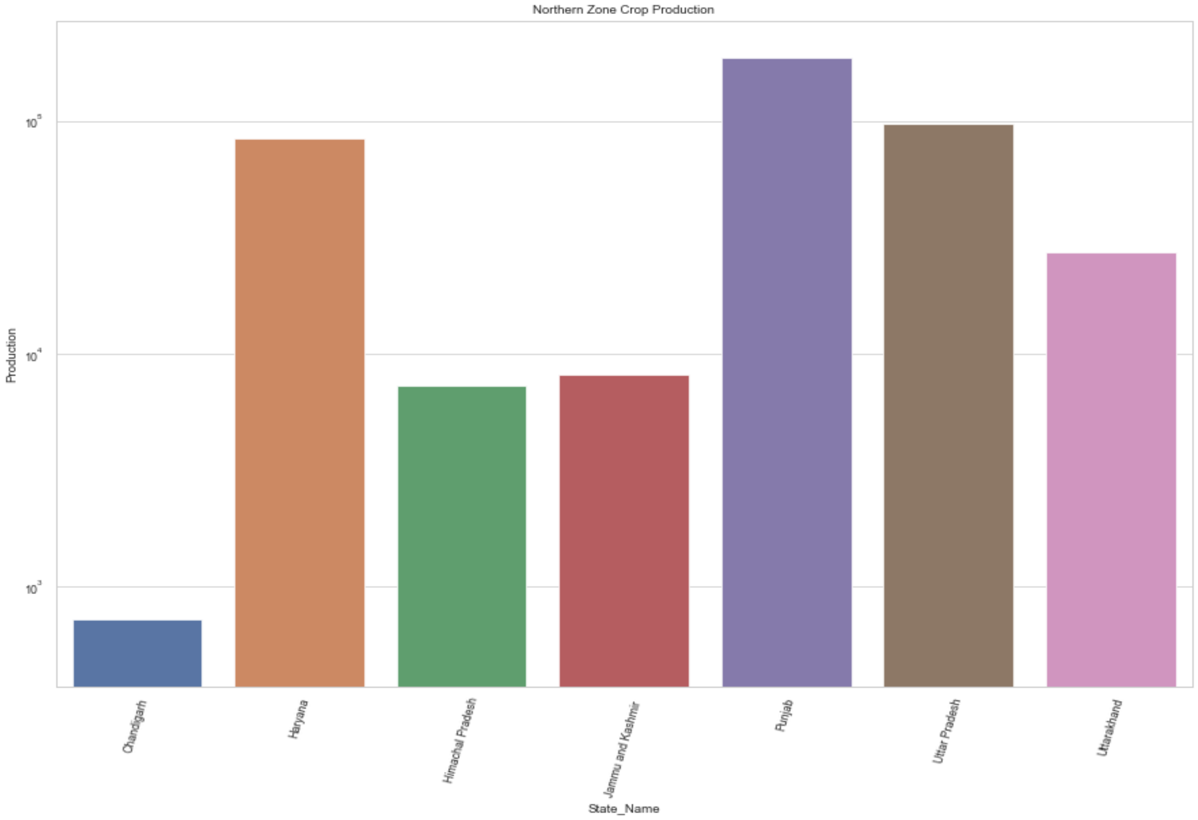

Figure 6: State wise crop production of North Zone of India

Table III: Top three most Statesfor Crop Production in zone of India

South	**East**	**West**	**North**
Kerala	West Bengal	Goa	Punjab
Andhra Pradesh	Bihar	Maharastra	Uttar Pradesh
Tamil Nadu	Odisha	Gujarat	Haryana

Next for each zone, found the top three most crops. These are represented from Figure 7 to Figure 10 and the outcome has stored in Table IV.

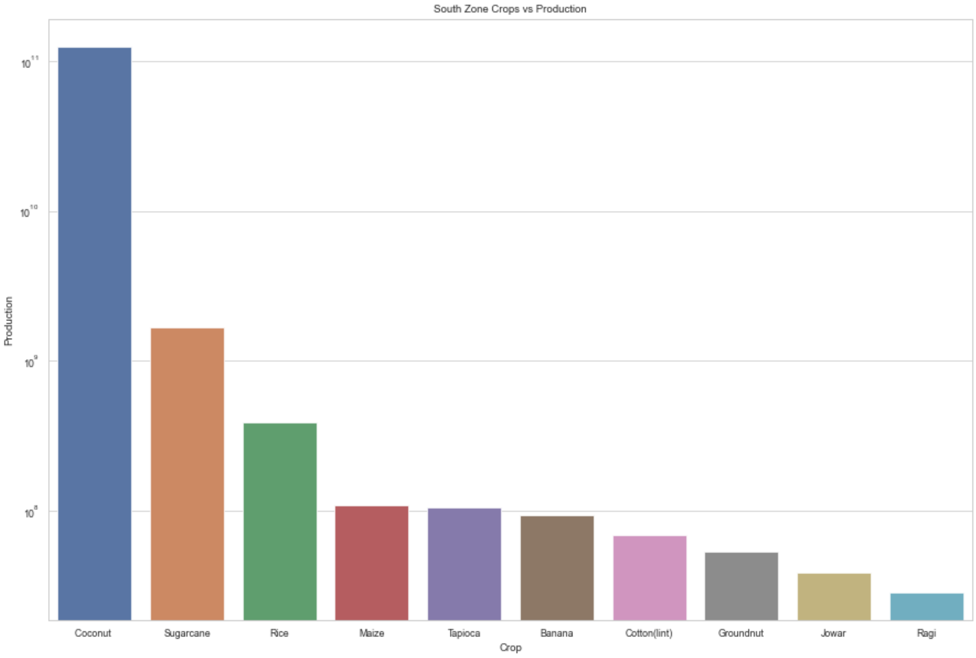

Figure 7: Crop production of South Zone of India

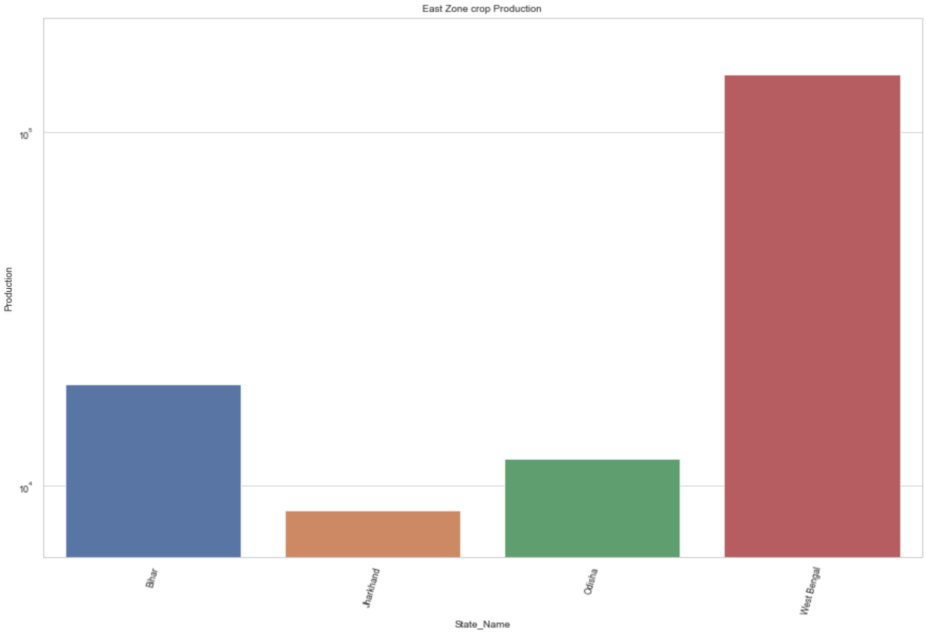

Figure 8: Crop production of East Zone of India

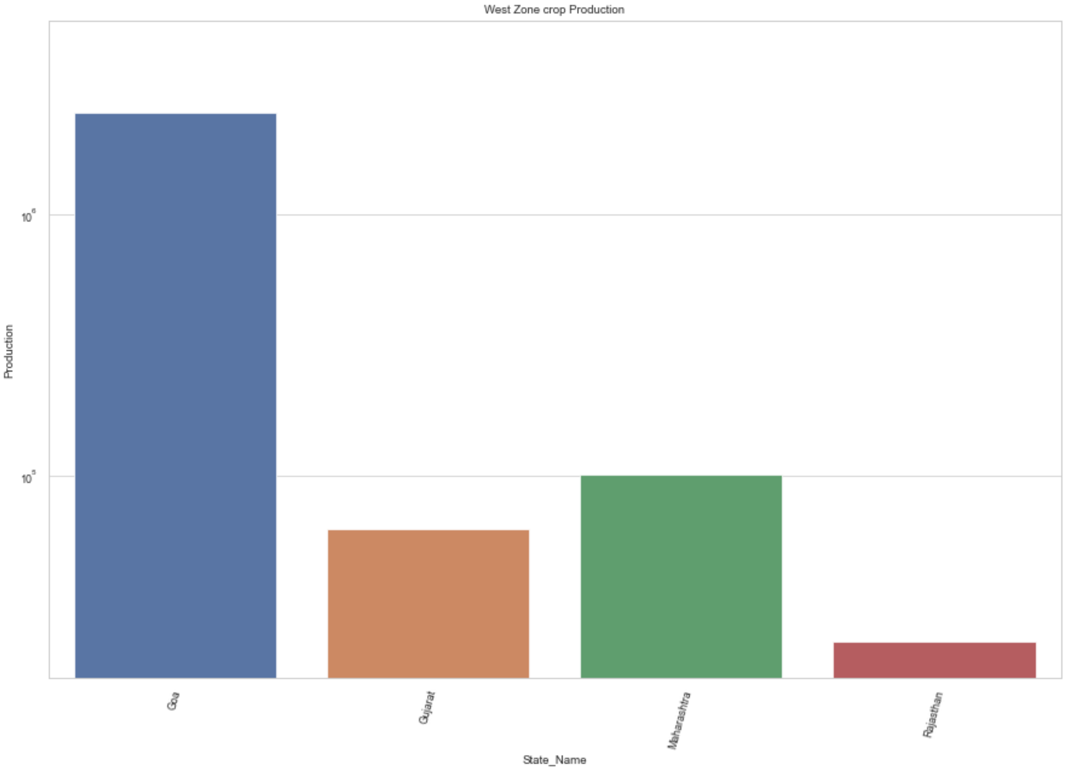

Figure 9: Crop production of West Zone of India

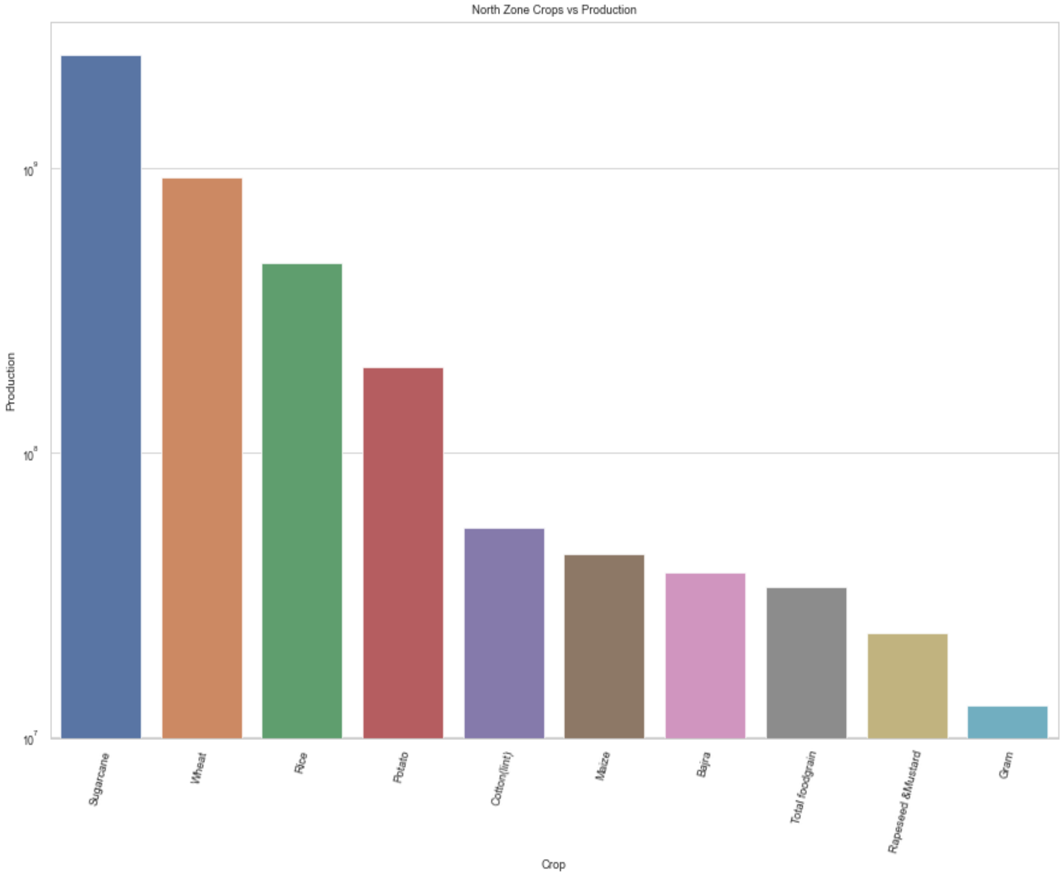

Figure 10: Crop production of North Zone of India

Table III: Top three most crops for each zone of India

South	East	West	North
Coconut	Coconut	Sugarcane	Sugarcane
Sugarcane	Rice	Coconut	Wheat
Rice	Rice	Cotton	Rice

From the outcomes, it has been found that Coconut, Rice, Sugarcane, Coconut, Cotton and wheat are the crops cultivated in various states and zones of India.

Conclusion and Future Work

In this work we carried out an experimental work to compare the analysis i.e. EDA ofcrop yield prediction using various accuracies over crop yield production of India. For this crop yield prediction, it has been found that Coconut, Rice, Sugarcane, Coconut, Cotton and wheat are the crops cultivated in various states and zones of India. The percentage of accuracy as well as prediction is highly determined by the data being utilized as input for prediction and ML models.

The accuracy of the prediction for crop can be predicted by using the varied ML model. In future work, planning for implementing the traditional ML models and hybrid prediction model to get the better and higher accuracy.

References

1. AnshalSavla, HimtanayaBhadada, ParulDhawan, Vatsa Joshi, Application of Machine Learning Techniques for Yield Prediction on Delineated Zones in Precision Agriculture, May 2015.
2. Beulah, R., 2019. A survey on different data mining techniques for crop yield prediction. Int. J. Comput. Sci. Eng. 7 (1), 738–744. https://doi.org/10.26438/ijcse/v7i1.738744.
3. D Ramesh, B Vishnu Vardhan, Analysis Of Crop Yield Prediction Using Data Mining Techniques, International Journal of Research in Engineering and Technology, Jan-2015.
4. Elavarasan, D., Vincent, D.R., Sharma, V., Zomaya, A.Y., Srinivasan, K., 2018. Forecasting yield by integrating agrarian factors and machine learning models: a survey. Comput.Electron. Agric. 155, 257–282. https://doi.org/10.1016/j.compag.2018.10.024.
5. GourHariSantra, Debahuti Mishra and Subhadra Mishra, Applications of Machine Learning Techniques in Agricultural Crop Production, Indian Journal of Science and Technology, October 2016.
6. Horie, T., Yajima, M., and Nakagawa, H. (1992). Yield forecasting. Agric. Syst. 40, 211–236. doi: 10.1016/0308-521X(92)90022-G.
7. Karan deep Kauri, Machine Learning: Applications in Indian Agriculture, International Journal of Advanced Research in Computer and Communication Engineering, April 2016.
8. Khaki, Saeed& Wang, Lizhi.(2019). Crop Yield Prediction Using Deep Neural Networks.Frontiers in Plant Science.10. 10.3389/fpls.2019.00621.
9. Klompenburg, Thomas &Kassahun, Ayalew&Catal, Cagatay. (2020). Crop yield prediction using machine learning: A systematic literature review. Computers and Electronics in Agriculture. 177. 105709. 10.1016/j.compag.2020.105709.
10. Lever, J., Krzywinski, M. & Altman, N. Principal component analysis.Nat Methods 14, 641–642 (2017). https://doi.org/10.1038/nmeth.4346
11. Li, B., Lecourt, J., Bishop, G., 2018. Advances in non-destructive early assessment of fruit ripeness towards defining optimal time of harvest and yield prediction—a review. Plants 7 (1). https://doi.org/10.3390/plants7010003.

12. Liakos, K.G., Busato, P., Moshou, D., Pearson, S., Bochtis, D., 2018. Machine learning in agriculture: a review. Sensors (Switzerland) 18 (8).https://doi.org/10.3390/s18082674.
13. Medar, R., Rajpurohit, V. S., &Shweta, S. (2019, March). Crop Yield Prediction using Machine Learning Techniques. In 2019 IEEE 5th International Conference for Convergence in Technology (I2CT) (pp. 1-5). IEEE.
14. Nishit Jain, Amit Kumar, SahilGarud, Vishal Pradhan, PrajaktaKulkarni, Crop Selection Method Based on Various Environmental Factors Using Machine Learning, Feb -2017.
15. Raorane A.A, Dr.Kulkarni R.V, Application of Data mining Tool to Crop Management System, January 2015.
16. SasmitaKumaraiNayak, Swati SucharitaBarik&MamataBeura. (2020).” Weather Forecasts Based on Rainfall Prediction Using Machine Learning Methodologies,” Adalya Journal 9 (6), Page No : 72 – 80. https://doi.org/10.37896/aj9.6/009
17. SasmitaKumaraiNayak, Swati SucharitaBarik, MamataBeura,” Analysis of Infectious Hepatitis Disease with High Accuracy Using Machine Learning Techniques,” TEST Engineering & Management 83 (Vol. 83: May/June 2020), 14294-14302.
18. SitiKhairunniza-Bejo, SamihahMustaffha and Wan Ishak Wan Ismail , Application of Artificial Neural Network in Predicting, Journal of Food Science and Engineering, January 20, 2014.
19. Sripada Swain, SasmitaKumariNayak, Swati SucharitaBarik. (2020).” A Review on Plant Leaf Diseases Detection and Classification Based on Machine Learning Models,” Muktshabd 9 (6), 5195-5205. DOI:09.0014.MSJ.2020.V9I6.0086781.105023
20. Tapas Ranjan Jena, Swati SucharitaBarik, SasmitaKumaraiNayak. (2020).” Electricity Consumption & Prediction using Machine Learning Models,” Muktshabd 9 (6), 2804-2818. DOI:09.0014.MSJ.2020.V9I6.0086781.104774