Sugarcane Production Modeling Using Machine Learning in Western Maharashtra

Chhaya Narvekar[1, *], Madhuri Rao[2]

[1]*Department of Information Technology, Xavier Institute Of Engineering , Mumbai , India*
[2]*Thadomal Shahani Engineering College , Mumbai, India*
[*]*Corresponding Author: chhaya.n@xavier.ac.in*

Abstract

Agriculture is the most important sector in the Indian economy. India is the world's second-largest producer of sugarcane. Study is undertaken at Shirol tehsil. Kolhapur district, Maharashtra state, India with the aim of modeling sugarcane production forecasting using supervised machine learning algorithms. Sugarcane is mostly cultivated crop in this area. We applied supervised machine learning for forecasting the productivity of sugarcane village wise based on the ten year's data about sugarcane production from the year 2010 to 2020. Sugarcane yield prediction accuracy is around 65%, which is only based on data provided by sugar factory.

Keywords: sugarcane, productivity, machine learning, forecasting.

1   Introduction

The Indian economy relies heavily on sugarcane cultivation. Sugar, as well as enterprises manufacturing alcohol, paper, chemicals, and animal feed, rely on it for raw materials. In India, sugarcane production is processed through a network of sugar mills, as well as various other businesses and backward and forward connections. The demand

for higher sugarcane production in India is increasing due to the multi-purpose usage of sugarcane in India and its byproducts in numerous sectors [1]. Despite rising urbanization around the world, agriculture remains the primary source of income for a huge percentage of the people. Although technological developments have resulted in more accurate weather predictions and increased yields, much work remains to be done to provide farmers with a taking into account local data so they can forecast yields. In the Maharashtra (India) region, the Sugarcane Cultivation Life Cycle (SCLC) spans around 12 months, with plantation starting at three separate seasons. Our method relies on past production data to train a supervised machine learning system and make sugarcane crop predictions.

Climate, production environments and agronomic aspects associated with agricultural management, such as variety selection, cane field age, fertilization, pest and disease control, and weed control, all influence sugarcane yield [2].

Description of study area - Shirol Taluka of Kolhapur district is gifted by the presence of natural irrigation potential on account of five major rivers i.e. Krishna,Panchaganga,Warana,Dudhganga and Vedganga [3]. Soil type here is alluvial. Normal rainfall is during June-October 1019.5mm. Top three crops cultivated are sugarcane 113.9('000 ha), Paddy Rainfed 113.8 ('000 ha) and Groundnut 57.4('000 ha) [4]. India is the world's second-largest producer of sugarcane after Brazil. Sugarcane is grown in all of India's states and at various times of the year. In this study, we propose supervised machine learning based crop yield forecasting model for sugarcane as a principal crop in study area. Crop analysis and agricultural production forecasting always relied on statistical models. Models are applied on ten years production data of the sugarcane. Three algorithms applied for sugarcane productivity prediction and five algorithms are applied for sugarcane yield prediction on ten years sugarcane production data from study area provided by Shree Dutta Sugar Factory, Shirol.

2   Research Methodology

Materials and Methods: Sugarcane is India's most important cash crop. It entails less risk, and farmers may be quite certain of a return even in difficult conditions. Sugarcane is first crop of Kolhapur district [4]. The sugarcane yield data, in tons of cane per hectare [5], originally available at the farmers and village gat number level. Ten years data from the sugar mill which includes farmer name, gat number village, date of sowing and season area of sugarcane cultivated and production. Knowing the size of the sugarcane harvest might assist industry members make better decision [2].

Table 1. Ten year sugarcane cultivation trend in study area

| Season-Year	Cultivated AREA	Total Sugarcane Production in Ton
2010-2011	14556.33	1344688.952
2011-2012	13032.36	1229240.511
2012-2013	10824.94	1196219.045
2013-2014	11139.9	1191862.504
2014-2015	337.8816667	67610.66034
2015-2016	11425.21	1294479.054
2016-2017	15058.3365	1224696.921
2017-2018	10524.99	1187021.203
2018-2019	12118.64	1212491.125
2019-2020	11637.48	1047024.887
2020-2021	11272.86	1192268.53
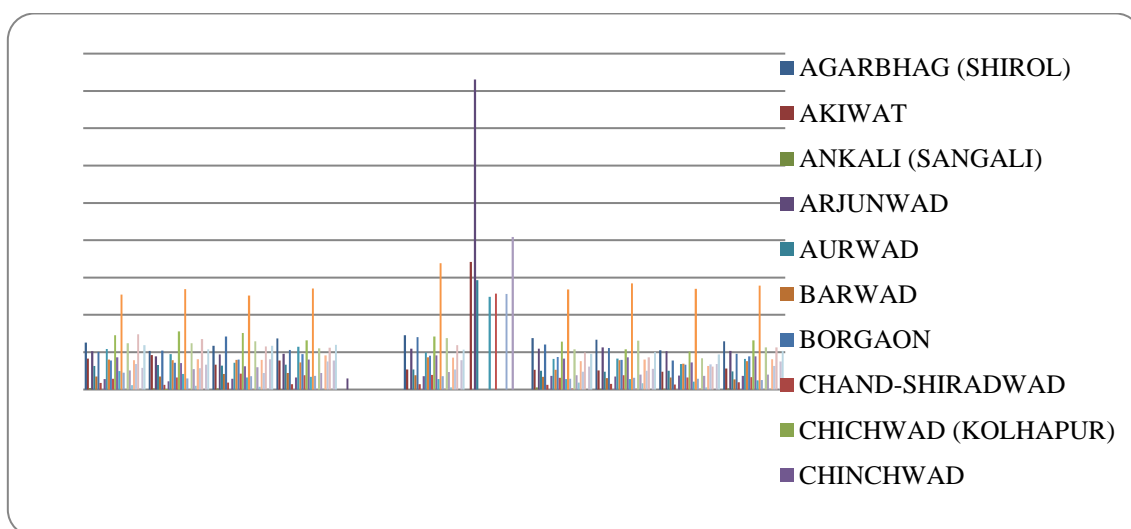
Figure 1. Village wise sugarcane production

From this we added column for productivity and village wise data created and applied machine learning algorithm for predicting the productivity of a particular village. Table 1, shows season wise sugarcane cultivated area and sugarcane production from study area. The model's predictor variables productivity of village is calculated on a yearly basis.

Regression analysis is a basic, technique for modeling the relationship between one or more independent or predictor variables and a dependent or response variable that we want to forecast, and it is one of the tools available in statistical analysis literature [5]. Sugarcane production in study area is summarized in Table 1.

Dataset description

Following Figure 2. shows the sample dataset which is recorded by sugar factory. Year wise production also visualised in Figure 3. For applying machine learning algorithm for yield some of the columns are removed which are less correlated. The features shown in figure 4 are used for training ML regression models after doing the pre-processing such as converting categorical variable in to numerical we get 57495 rows x 11 columns . Dataset is further divided into 80% training and 20 % testing . Performance of model discussed in results section.

	SEASON	VILLAGE_NAME	AREA	WATER_SUPPLY_TYPE	CAME_VARIETY	compute_0012	MATURITY_TYPE	ACTUAL_TON
0	2014-2015	ARJUNWAD	0.40	RIVER	CO-0265	ADASALI	LATE	58.424
1	2014-2015	ARJUNWAD	0.10	RIVER	CO-86032	ADASALI	MIDDLE	15.751
2	2014-2015	ARJUNWAD	0.15	RIVER	CO-0265	ADASALI	LATE	25.740
3	2014-2015	ARJUNWAD	0.40	RIVER	CO-0265	ADASALI	LATE	76.116
4	2014-2015	ARJUNWAD	0.80	RIVER	CO-0265	ADASALI	LATE	129.497

Figure 2. Sample Recorded Data
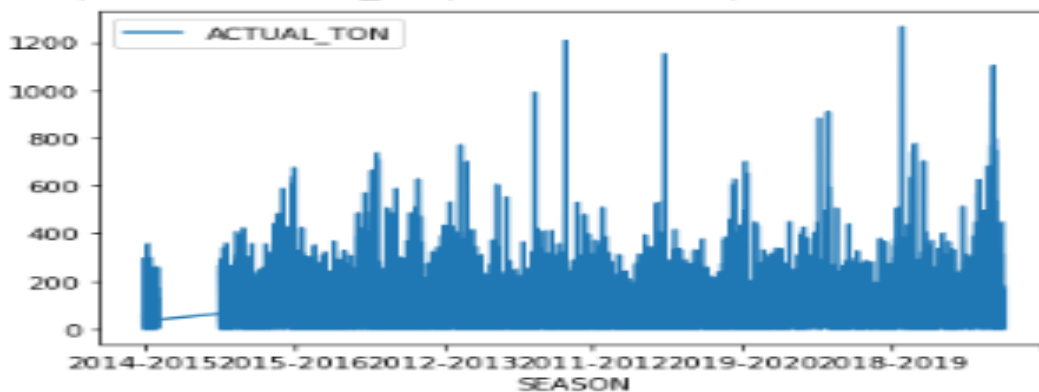
Figure 3. Yearly Production

Figure 4. Features Used for Modeling

From the data provided another dataset created which is village wise yearly cultivated area and village wise sugarcane production and productivity of each village calculated per unit area production and machine learning models are trained and tested on this created dataset as well for forecasting productivity of particular village. Productivity compared with national level productivity and state level productivity to get further insights. In both cases climatic, nutrient supply, soil fertility status such parameters are not taken into consideration, which can taken and accuracy would be improved.

3   Results and Discussion

Crop forecasting is the science of estimating crop yields and production ahead of time, usually a couple of months ahead of time. A crucial part of crop production

forecasting is defining the time horizon in terms of time series forecasting approaches. This study included three algorithm for productivity prediction   random forest (RF), boosting (GBM), and XGboost which are the most commonly used for agricultural modeling[6]. We tried five different algorithms for modeling yield [4] performance is as shown in Table 2. Performance is not great because there are extrinsic parameters as well which impact on production of crop such as climate, rainfall, soil fertility , management skill and so on which are not considered in the current study.

Table 2. Sugarcane Yield Prediction Model Performance

Machine Learning Algorithm	Accuracy
Linear Regression	62%
Random Forest	65%
XGboost	66%
Gradient Boost	63%
Decision Tree	63%

For sugarcane productivity prediction only village wise area cultivated and production of sugarcane for that particular season is used and target variable is productivity. In this modeling climate data , rainfall , soil quality not considered , parameters used are how much area cultivated, which type of breed, when it is planted, what type of water supply and when crop is taken. Average sugarcane productivity of India is 70-80, average sugarcane productivity of Maharashtra 80.72, [R20], whereas average productivity of the study area is 95. Random forest repressor gives 65% accuracy and other two XGBoost and gradient boosting gave 66% accuracy. When opposed to using a single data model to predict a response, using many model can improve the robustness and accuracy of predictions.

4   Conclusion

The goal of this study was to see if a machine learning approach could provide fresh insights about sugarcane productivity in western Maharashtra. Predicting crop

production may help sugar mills to boost industry revenues by implementing more effective and focused forward selling tactics and logistics planning. The methodology described in this research can readily be applied to other sugarcane-growing regions and agricultural businesses around the world to improve agricultural methods. Sugarcane productivity prediction results demonstrated that the prediction accuracy of the machine learning algorithm is quite promising.

Acknowledgements

References

[1] P. Mishra, M. A. G. A. Khatib, I. Sardar, J. Mohammed, K. Karakaya, A. Dash, M. Ray, L. Narsimhaiah, A. Dubey, "Modeling and Forecasting of Sugarcane Production in India", *Sugar tech,* **23**(6), 1317-1324, 2021.

[2] L. A. Monteiro and P. C. Sentelhas, "Sugarcane yield gap: can it be determined at national level with a simple agrometeorological model?", *Crop and Pasture Science*, **68**(3), 272-284, 2017.

[3] I. MAHARASHTRA CELL, "Agriculture Contingency Plan for District: KOLHAPUR", ICAR_CRIDA_NICRA, 2019.

[4] Y. Everingham, J. Sexton, D. Skocaj, G. Inman-Bamber, "Accurate prediction of sugarcane yield using a random forest algorithm". *Agron. Sustain. Dev.* **36**. 27. Springer Verlag/EDP Sciences/INRA. 2016.

[5] R. G. Hammer, P. C. Sentelhas, J. C. Mariano, "Sugarcane yield prediction through data mining and crop simulation models". *Sugar Tech*, **22**(2), 216-225. 2020

[6] R. S. Kodeeshwari and K. T. Ilakkiya, "Different types of data mining techniques used in agriculture-a survey". *International Journal of Advanced Engineering Research and Science*, **4**(6), 237191. 2017.

[7] Shree Datta Shetkari S.S.K. Ltd., Shirol. Available at: http://dattasugar.co.in/