Analysis of Geo-Economic Factors of Production of Agriculture in India

***Abstract*—Uncertain yield of crops is one of the major problems the agricultural sector faces today. We attempt to produce a dataframe to predict the productivity and production of major crops of India and establish parameters which will be useful for doing so. The previous work focuses on particular states or using very few parameters whereas we aim to analyze all parameters which might affect production thus creating an exhaustive list and a framework for further data. In particular we have focused on geo-economic parameters and tried to encode them efficiently and eliminate redundancies as well as determine their role in this sector.**

I. Introduction

The prices of crops and vegetables are extremely volatile and the fluctuations are difficult to predict due to the numerous factors which affect it. In a country like India, where over 60 percent of the country depends on agriculture as its means of subsistence, it becomes especially important that we solve this issue. Agriculture has been the sector of paramount importance as it feeds the country population along with contributing to the GDP. The process of collection of geographical data has been going on for more than 30 years in India and there is enough accumulation of data now, to drive the Machine learning models for prediction. However this data is highly scattered across different government websites. Through this research we have attempted establish parameters to predict the yield of the crop district-wise in India. Due to the high correlation between prices and yield as shown by (Bruce et al, 2012), predicting prices boils down to predicting yield. Moreover, data for yield is much more widely available than prices, allowing us to use Big Data and Machine Learning in a much more powerful way. The objective of the research is to provide a foundation of data for learning agents that can help the farmers to take decisions that can make their farming more efficient and profitable as compared to the traditional methods. Moreover, this can also help the government make timely import and export decisions based on accurate predictions and also set the Minimum Selling Price as well as serve as a guideline for future data collection and organization. This will also help farmers roughly estimate their production and price that they can get,thus giving them a bargaining point to establish prices in contract farming beforehand(A problem of significant interest currently) The attempts to solve this problem mainly revolve around the prediction algorithm and techniques whereas our focus is on the parameters and the data which can help improve these models and make them more robust by using a large dataset.We prove the superiority of our dataset by comparing results on standard learning algorithms used for predicting agricultural yield

II. Related Works

Predicting the yield of crops is one of the most complex problems, owing to the gargantuan amount of data and numerous factors. In order to tackle this problem, we need a detailed study of already existing methods and cutting-edge algorithms designed for this purpose. These are some of the most recent research papers on using Machine Learning techniques to approach such problems

A. Random Forest Algorithm

Random forest is a powerful supervised machine learning algorithm capable of performing , regression tasks, that operate by constructing a multitude of decision trees at training time and giving the output as the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees (Priya, Muthaiah, Balamurugan, 2018)

B. K Nearest neighbors Algorithm

K nearest neighbor is a supervised machine learning classification algorithm. This algorithm basically searches for similarities in the entire dataset. Based on the most similar or nearest values the outcome is predicted. (Rachana, Rashmi, Shravani, Shruthi, Seema, 2019)

C. Recurrent Neural Network

RNNs are used for tasks involving sequential data to capture their time dependencies hence making them perfect for this problem where lot of sequential data like temperature and rainfall is involved (S. Kulkarni et al 2020). RNNs keep the history of all the past elements of a sequence in their hidden units called a state vector and use this information as they process input sequence one element at a time (LeCun, Bengio, Hinton, 2015).

D. Convolutional Neural Network

CNNs process data with multiple arrays format such as one-dimensional data (signals and sequences), two-dimensional data (images), and three-dimensional data (videos). One dimensional CNNs can be used to process temporal data. (Saeed, Lizhi, Sotirious, 2020).

III. Datasets

A big challenge in current machine learning research is mining all the relevant data and converting it to the required format for the models. We managed to collect data across various seasons, or many crops, across several districts for the years 1997 to 2015

A. Temperature data (maximum and minimum temperature)

The temperature data was obtained at IMD. The data was provided as a (1.0X1.0) GRD file so we Wrote a C++ program to extract it and used the latitudes and longitudes data of each district so as to get the data district wise. The data was available per day so we calculated the monthly average. We also calculated the monthly average across the years. The temperature values are supposed to give an idea about the whether conditions prevailing in a particular region at a particular point in time. Weather conditions are an important factor for the growth of crops. Hence, they were required to predict the yield.

B. Rainfall

The rainfall data was obtained at IMD. This data was available as a (0.25X0.25) GRD file. We again used C++ to extract it. This data was also available day wise so we calculated the monthly average. We also calculated the monthly average across the years. Like the temperature data, the rainfall data also plays an important role in determining the weather conditions of the region. In addition to that, they also give an idea about the natural irrigation available for the crops. Different crops have different water requirements. Hence this was also an important factor while predicting the yield.

C. Latitude and Longitude of districts

For the data about Latitude's and Longitude's of each district in our data frame we scraped google using selenium.

D. State wise Population

We got the Population data from RBI website. The website provided us with population of different states of India.

E. Per capita NDP/GDP/Agricultural GDP/Literacy Rate

The RBI website also provided us with these datasets. The datasets were also present state wise.

F. Area under Cultivation and production data

We found the area under Cultivation and production data on the government website data.gov.in. The dataset was present district-wise.

G. District wise area under cultivation/District Population/District yield

We got this data from the website of census of India.

H. State Population Density

We found this from the census of India.

IV. Analysis Pipeline

A. Data Cleaning

1) Deduplication: -Deleting duplicate rows was particularly tough and needed to be done before merging of the data for each dataset as well as after merging the data(generated due to uneven granularity)

2) Null Value treatment: -Due to the diverse nature of data corresponding to varied sources, the rows were bound to have null values generated especially from GRD files as it had daily temperature and rainfall values. To deal with this problem the data was filled by taking averages of closest dates having data points in that year, as well as averaging it across different years. In case of missing data in the to be predicted columns,the rows had to be deleted as no clear method was available to fill the null values.

B. Merging the dataset

After collecting data for different parameters primarily governing Agriculture,the next step is to merge it to create a data frame.As data was collected from various different sources, a prime challenge was selecting granularity by keeping low enough to encompass all data as well as high enough to have enough rows of data to make accurate predictions. Facing this trade-off and keeping in mind that crops are generally produced seasonally, we decided to predict productivity for a particular crop in a particular district in a particular season.The columns of geographical significance correspond to average rainfall in each month as well as average min and average max temperature in that month.Along with this economic parameters such as GDP of state,NDP of nation, percentage of GDP corresponding to agriculture,etc were taken

C. Creating the final Dataframe

1) Correlation Coefficient: Correlation matrix was computed for the given numerical columns of the data and a heatmap was created for the same. This showed high correlation between the geographic factors like temperature and rainfall. Then the two were analyzed separately. Due to high correlation in rainfall data across years, a single value to encode rainfall across a year was enough as correlations exceeded 0.9 for all months. The temperature data showed comparatively less correlation but there was a clear pattern of correlation across different seasons. The temperature data was thus compressed accordingly. This resulted in a significant size reduction having much of the information still intact.There was however much lesser correlation between the economic factors and all of them were chosen for making predictions. It is also interesting to note how little correlation the production as well as productivity have with the rest of the factors thus creating a need for a complex non linear model for their prediction

2) One Hot Encoding: All the categorical data type columns were encoded using one hot vectors.This included States,districts,seasons as well as crop.A possible way to reduce size of One hot encoding of districts was to use their latitude and longitude values. This method however causes a massive loss in information as a districts productivity of a particular crop isn't only dependant on its geographical location but also on much more complex factors which two degrees of freedom cannot encode.

3) Two possible data frames: The productivity of different crops is majorly independent of others thus creating a need for a different model.This however results in fewer rows of data and weaker learning agent as data gets split for different crops. Another point in favour of joint dataset is that even though growth of crops is majorly independent of others,there do exist some common factors like irrigation,fertility of soil , state taxes which tend to affect all crops in general.The only way to conclusively find which method is better is to make prediction using both to see which gives better results

4) Choosing the Crops and Districts: The crops were chosen which are primarily grown in India by comparing their production across India.Similarly the districts were chosen for a particular crop corresponding to the top areas of production.This was done to reduced outliers and skewed data points.

D. Prediction

In the model part, we used 5 different regression models namely, Linear Regressor, XGB Regressor, Random Forest Regressor, SVM Regressor with Gaussian Kernel and MLP Regressor. These 5 models were trained on datasets corresponding to 6 different crops having the highest production values all over the country. We used a train-test split ratio of 70:30. We used all variables except Production and productivity to predict the target variable. We ran the model twice for each dataset, once with target variable as "Production" and another time with target variable as "Productivity".

Further we used 4 metrics to test the correctness of our predictions. These were, the model score (inbuilt), Correlation Coefficient Loss, MAPE Loss, MSE loss.

Further we also standardised our dataset. This was to avoid skewed data. This can be understood by an easy example. Suppose there is a single value of a target variable which is extremely high. Obviously, the model isn't trained to predict it and moreover, that value can't be trusted also. We don't want our model to train as per those values. However, when we calculate the error corresponding to that case, the value would definitely shoot up, giving a wrong implication about the model.

All this part was done very efficiently. The whole process was divided, at appropriate places, into functions which were called as and when required. This made the task very easy and making changes was quite efficient. The functions were made such that they could fit well with our needs. They provided us with various options such as choosing whether to predict production or productivity, using standardisation or not and many more. This made the work a lot cleaner and easier to understand. The modularity of the code helped us compare our results with the various options available.

V. Results

The following are our results from the code and model we ran.

Production results for Coconut with standarisation

Model	Train Score	Test Score	Corr Coeff Loss	Corr Coeff Loss Test	MAPE Loss Train	MAPE Loss Test	MSE Loss Train	MSE Loss Test
Linear Regression	0.8457	0.7664	0.84571	0.766858	0.1806	0.189	0.40004	0.4612
Randon Forest Regressor	0.9863	0.8434	3.65E-34	7.14E-32	0.0305	0.0901	0.11935	0.3776
XGB Regressor	0.9847	0.8824	6.50E-33	1.14E-33	0.0552	0.1012	0.12605	0.3272
SVM Regressor with Gaussian	-0.056	-0.036	2.78E-32	1.76E-32	0.4059	0.3526	1.04643	0.9713
MLP Regressor	-1E+07	-1E+07	9.05E-32	2.31E-33	2345.4	2353.1	3252.86	3256.9

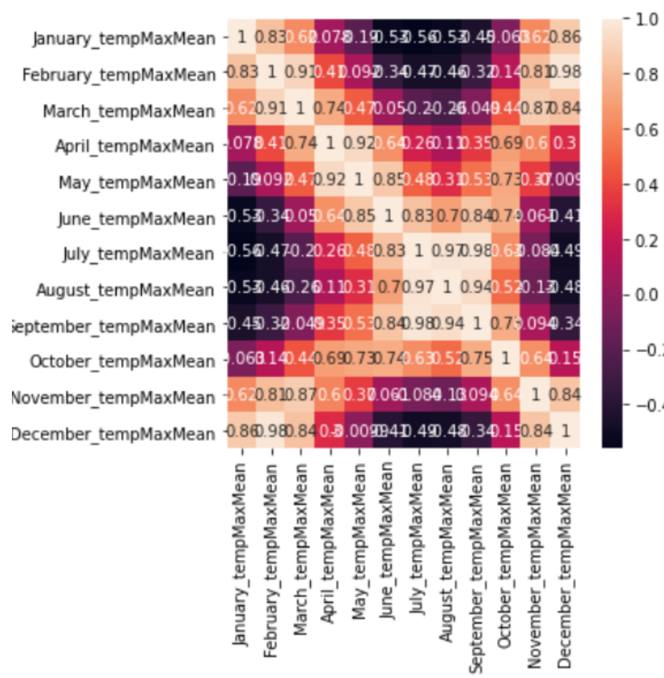

Fig. 1. Coorelation between maximum temperature

VI. Discussion

From the results we obtained, shown in the above section (see appendix for detailed results), we observed that MLP Regressor didn't provide convincing results. This is in support with the observation that there isn't much correlation between production (or productivity) and other input variables. Hence linear models tend to perform worse. The methodology of dividing the datasets cropwise turned out to be helpful as training models on complete dataset turned out to be very time consuming and hence not very scalable approach. Dividing cropwise is also logically sound because there seems to be very little relation between the yields and over all requirements of two different crops. Combining the temperature and rainfall data was also beneficial since they were highly correlated and that would not have lead the model into any other meaningful direction. Standardising the data was a major step and it improved the results and reduced the train time many folds.

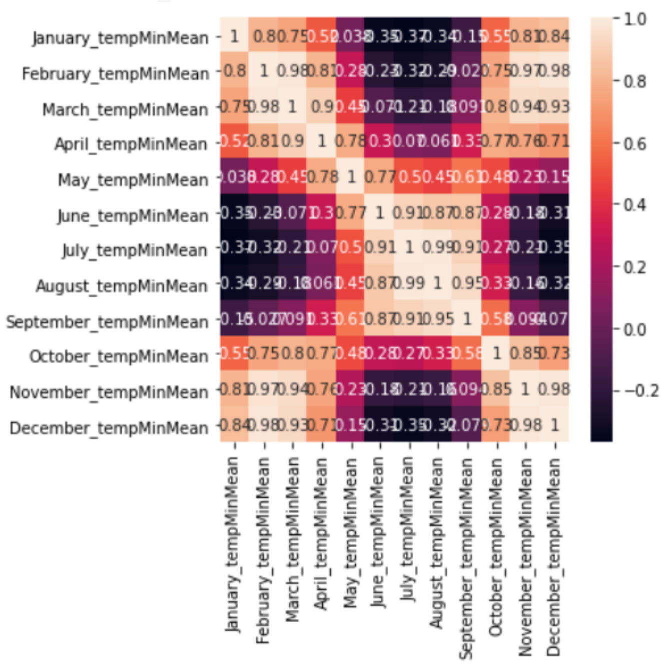

Fig. 2. Coorelation between minimum temperature

From our analysis, models and their results, we concluded that if more sophisticated models were trained on our datasets which are made in such a way to avoid the little over-fitting which had remained in our approach, then our idea would become really very helpful in predicting the yield of crops. This could be used by government agencies for formulating policies and schemes which would be beneficial to the farmers. This approach is realisable because all the variables which we have used such as temperature, rainfall, etc are predicted with very high accuracies in today's date. We could use this in our model to predict the yields of various crops in different regions.

Our idea can be extended to making models which could be trained on regionwise data.

VII. Acknowledgment

We would like to thank the Indian Government websites like IMD,RBI and Census of India for maintaining records of agricultural and economic significance and providing high frequency and accurate data for free accompanied by the code to extract them.We would also like to thank the course instructors of Programming for Data Science (DS 203) for giving us this opportunity and the teaching assistants for providing invaluable feedback and help throughout this course

References

[1] Borovykh, A., Bohte, S., Oosterlee, C. W. (2017). Conditional time series forecasting with convolutional neural networks.

[2] Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory. Neural Comput.

[3] Hsu, Y. K., Ting, H. K., Chi, H. C., Pei, Y. T. (2016). Accuracy Analysis Mechanism for Agriculture Data using the Ensemble Neural Network Method.

[4] Ince, T., Kiranyaz, S., Eren, L., Askar, M., Gabbouj, M. (2016). Real-time motor fault detection by 1-d convolutional neural networks. IEEE.

[5] Kiranyaz, S., Ince, T., Abdeljaber, O., Avci, O., Gabbouj, M. (2019). 1-d convolutional neural networks for signal processing applications. IEEE ICASSP.

[6] LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep Learning. Nature 521, 426.

[7] Ng, A. Y. (2004). Feature selection, L1 vs. L2 regularization, and rotational invariance. Proceedings of the Twenty-first International Conference on Machine learning.

[8] Pham, V., Bluche, T., Kermorvant, C., Louradour, J. (2014). Dropout improves recurrent neural networks for handwriting recognition. 14th International Conference on Frontiers in Handwriting Recognition.

[9] Priya, P., Muthaiah, U., Balamurugan, M. (2018). PREDICTING YIELD OF THE CROP USING MACHINE LEARNING ALGORITHM. IJESRT, 1-7.

[10] Rachana, P. S., Rashmi, G., Shravani, D., Shruthi, N., Seema, K. R. (2019). CROP PRICE FORECASTING SYSTEM USING SUPERVISED MACHINE LEARNING ALGORITHMS. IRJET, 1-3.

[11] Saeed, K., Lizhi, W., Sotirious, V. A. (2020, January 24). A CNN-RNN Framework for Crop Yield Prediction. Frontiers in Plant Science, 1-14.

[12] Hornik, K., Stinchcombe, M., White, H. (1989). Multilayer feedforward networks are universal approximators. Neural Networks 2, 359-356.

[13] Kingma, D. P., Ba, J. (2014). Adam: A method for stochastic optimization.

[14] Sherstinsky, A. (2018). Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network.

[15] Glorot, X., Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics.

[16] Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep Learning Vol. 1. Cambridge: MIT Press.

Appendix

Production results for Sugarcane with standarisation								
Model	Train Score	Test Score	Corr Coeff Loss	Corr Coeff Loss Test	MAPE Loss Train	MAPE Loss Test	MSE Loss Train	MSE Loss Test
Linear Regression	0.8577	0.8771	0.85769	0.877765	0.0947	0.0997	0.38304	0.3377
Randon Forest Regressor	0.9929	0.959	1.15E-32	1.60E-34	0.0174	0.0436	0.08551	0.195
XGB Regressor	0.9815	0.9436	9.24E-34	1.94E-32	0.0469	0.0627	0.13807	0.2288
SVM Regressor with Gaussian	0.0075	0.0039	1.07E-32	7.02E-32	0.3538	0.3653	1.01155	0.9613
MLP Regressor	-4E+06	-4E+06	7.15E-33	7.34E-32	1274.4	1275.9	1910.21	1893.2

Production results for Rice with standarisation								
Model	Train Score	Test Score	Corr Coeff Loss	Corr Coeff Loss Test	MAPE Loss Train	MAPE Loss Test	MSE Loss Train	MSE Loss Test
Linear Regression	0.8983	0.9056	0.89826	0.906061	0.1657	0.1697	0.3184	0.3085
Randon Forest Regressor	0.9945	0.9687	5.78E-32	4.31E-32	0.0314	0.0785	0.07435	0.1778
XGB Regressor	0.9482	0.9393	3.55E-32	1.23E-31	0.127	0.1368	0.22723	0.2473
SVM Regressor with Gaussian	0.0138	0.0093	1.52E-32	2.94E-32	0.5274	0.532	0.99133	0.9994
MLP Regressor	-4E+06	-4E+06	1.23E-31	4.19E-33	1452.7	1458.4	1993.48	1967.6

. ..

Production results for Wheat with standarisation								
Model	Train Score	Test Score	Corr Coeff Loss	Corr Coeff Loss Test	MAPE Loss Train	MAPE Loss Test	MSE Loss Train	MSE Loss Test
Linear Regression	0.9751	0.9769	0.97513	0.977005	0.0903	0.0955	0.15502	0.1576
Randon Forest Regressor	0.9973	0.9839	1.18E-31	1.31E-31	0.0255	0.0683	0.05084	0.1316
XGB Regressor	0.9821	0.979	8.10E-33	6.30E-32	0.0819	0.0943	0.13137	0.1505
SVM Regressor with Gaussian	0.2251	0.2275	1.25E-32	1.41E-32	0.4512	0.4704	0.8653	0.9125
MLP Regressor	-1E+07	-1E+07	6.57E-32	1.36E-32	2170.9	2149.4	3208.46	3205.1

Production results for Cotton with standarisation								
Model	Train Score	Test Score	Corr Coeff Loss	Corr Coeff Loss Test	MAPE Loss Train	MAPE Loss Test	MSE Loss Train	MSE Loss Test
Linear Regression	0.8167	0.7713	0.81673	0.777105	0.2033	0.219	0.44053	0.444
Randon Forest Regressor	0.9872	0.8984	1.12E-32	1.27E-31	0.039	0.1008	0.11632	0.2959
XGB Regressor	0.9612	0.8773	1.37E-33	1.47E-32	0.0884	0.1231	0.2026	0.3252
SVM Regressor with Gaussian	-0.031	-0.006	3.60E-35	3.22E-33	0.4164	0.3867	1.04498	0.9311
MLP Regressor	-2E+06	-2E+06	7.01E-33	1.55E-33	823.03	827.7	1383.15	1411.1

Production results for Potato with standarisation								
Model	Train Score	Test Score	Corr Coeff Loss	Corr Coeff Loss Test	MAPE Loss Train	MAPE Loss Test	MSE Loss Train	MSE Loss Test
Linear Regression	0.9402	0.9265	0.94015	0.926572	0.0856	0.099	0.23572	0.2928
Randon Forest Regressor	0.9942	0.9535	8.78E-34	5.76E-32	0.0211	0.0606	0.07307	0.2328
XGB Regressor	0.9856	0.9477	6.89E-32	1.70E-34	0.0524	0.0736	0.11543	0.2469
SVM Regressor with Gaussian	0.108	0.0891	1.53E-33	3.96E-32	0.3022	0.3261	0.91002	1.0308
MLP Regressor	-5E+05	-4E+05	1.55E-31	2.29E-31	473.24	496.89	663.898	700.5

Productivity results for Coconut with standarisation								
Model	Train Score	Test Score	Corr Coeff Loss Train	Corr Coeff Loss Test	MAPE Loss Train	MAPE Loss Test	MSE Loss Train	MSE Loss Test
Linear Regression	0.6807	0.4828	0.68069	0.490644	0.3854	0.445	0.5571	0.7413
Randon Forest	0.9791	0.825	1.32E-32	1.53E-32	0.0677	0.1581	0.14259	0.4312
XGB Regressor	0.8962	0.7505	2.63E-32	5.70E-33	0.1923	0.2453	0.31767	0.5149
SVM Regressor with	0.0586	0.0796	4.84E-33	4.08E-34	0.6142	0.5882	0.95659	0.9889
MLP Regressor	-1E+08	-1E+08	5.62E-32	2.52E-33	7288	7410.2	11282.1	11383

Productivity results for Sugarcane with standarisation								
Model	Train Score	Test Score	Corr Coeff Loss Train	Corr Coeff Loss Test	MAPE Loss Train	MAPE Loss Test	MSE Loss Train	MSE Loss Test
Linear Regression	0.1298	0.0138	0.12983	0.031277	0.1729	0.1801	0.92059	1.0228
Randon Forest	0.9967	0.9828	2.66E-32	3.16E-32	0.0038	0.0096	0.05655	0.1349
XGB Regressor	0.9999	0.9866	2.60E-33	2.42E-32	0.0042	0.0089	0.00769	0.1193
SVM Regressor with	-0.001	-7E-04	2.11E-33	1.61E-32	0.1136	0.1178	0.98742	1.0303
MLP Regressor	-1E+07	-1E+07	4.26E-32	1.49E-32	2310.5	2401.1	3248.03	3322.2

Productivity results for Wheat with standarisation								
Model	Train Score	Test Score	Corr Coeff Loss Train	Corr Coeff Loss Test	MAPE Loss Train	MAPE Loss Test	MSE Loss Train	MSE Loss Test
Linear Regression	0.8812	0.8757	0.88116	0.875697	0.2437	0.2613	0.34216	0.3586
Randon Forest	0.9877	0.9233	1.04E-32	1.67E-32	0.0751	0.1993	0.11002	0.2817
XGB Regressor	0.8611	0.8571	2.89E-34	3.09E-31	0.2831	0.2998	0.36987	0.3845
SVM Regressor with	0.1741	0.1763	1.71E-33	6.99E-33	0.6536	0.6618	0.90198	0.9232
MLP Regressor	-4E+07	-3E+07	2.61E-32	2.88E-32	4190.9	4055.3	5894.55	5645.5

Productivity results for Potato with standarisation								
Model	Train Score	Test Score	Corr Coeff Loss Train	Corr Coeff Loss Test	MAPE Loss Train	MAPE Loss Test	MSE Loss Train	MSE Loss Test
Linear Regression	0.682	0.6466	0.68202	0.654694	0.3021	0.3243	0.5735	0.57
Randon Forest	0.953	0.5518	1.10E-31	1.25E-32	0.087	0.2475	0.22038	0.642
XGB Regressor	0.8085	0.3722	3.43E-32	8.21E-34	0.3146	0.3448	0.44504	0.7598
SVM Regressor	0.2871	0.3197	2.70E-32	1.16E-32	0.5505	0.5493	0.85872	0.7908
MLP Regressor	-8E+05	-1E+06	5.31E-33	2.96E-32	626.63	647.29	930.303	983.5

Productivity results for Cotton with standarisation								
Model	Train Score	Test Score	Corr Coeff Loss Train	Corr Coeff Loss Test	MAPE Loss Train	MAPE Loss Test	MSE Loss Train	MSE Loss Test
Linear Regression	0.1887	0.0505	0.18871	0.051866	0.2186	0.2883	0.68974	1.3654
Randon Forest	0.9489	0.5839	2.41E-32	4.02E-34	0.0228	0.0903	0.17312	0.9039
XGB Regressor	0.9813	0.6398	6.61E-33	6.30E-32	0.0606	0.1031	0.10474	0.8409
SVM Regressor	-0.002	-0.003	1.72E-33	6.32E-32	0.1351	0.1784	0.76668	1.4031
MLP Regressor	-9E+06	-3E+06	6.79E-33	1.13E-32	1650.2	1708.1	2343.92	2424.5

Productivity results for Rice with standarisation								
Model	Train Score	Test Score	Corr Coeff Loss Train	Corr Coeff Loss Test	MAPE Loss Train	MAPE Loss Test	MSE Loss Train	MSE Loss Test
Linear Regression	0.1515	0.1878	0.15155	0.42771	0.1793	0.1848	1.07127	0.3794
Randon Forest	0.7858	-4.871	5.39E-32	1.08E-31	0.0482	0.1313	0.5383	1.0202
XGB Regressor	0.9465	-11.46	2.93E-33	1.54E-34	0.1797	0.2072	0.26909	1.4864
SVM Regressor with	0.0208	0.0896	1.00E-32	7.72E-32	0.3147	0.3043	1.15089	0.4017
MLP Regressor	-1E+06	-8E+06	4.58E-33	1.82E-32	851.14	843.35	1225.8	1219.9

Productivity results for Wheat without standarisation								
Model	Train Score	Test Score	Corr Coeff Loss Train	Corr Coeff Loss Test	MAPE Loss Train	MAPE Loss Test	MSE Loss Train	MSE Loss Test
Linear Regression	0.8812	0.8757	0.88116	0.875697	0.2573	0.2759	0.36128	0.3787
Randon Forest	0.9881	0.9236	0.98881	0.924734	0.0791	0.2113	0.11445	0.2968
XGB Regressor	0.8625	0.8566	0.86575	0.860653	0.2973	0.3182	0.38865	0.4067
SVM Regressor with	0.1731	0.1753	0.20032	0.202249	0.6905	0.6992	0.95297	0.9754
MLP Regressor	-1E+06	-1E+06	0.0015	0.000818	815.7	822.98	1221	1207.8

Productivity results for Coconut without standarisation								
Model	Train Score	Test Score	Corr Coeff Loss Train	Corr Coeff Loss Test	MAPE Loss Train	MAPE Loss Test	MSE Loss Train	MSE Loss Test
Linear Regression	0.6807	0.4828	0.68069	0.490643	1980	2286.5	2862.36	3808.7
Randon Forest	0.9792	0.8194	0.98034	0.82151	332.77	808.22	730.626	2250.4
XGB Regressor	0.9012	0.7456	0.91004	0.74835	974.11	1254.7	1592.11	2671.3
SVM Regressor with	-0.526	-0.4	0.03398	0.032166	3703	3383.7	6256.5	6267
MLP Regressor	-4.68	-4.541	0.00054	0.01368	9866.9	10405	12072.8	12466

Productivity results for Cotton without standarisation								
Model	Train Score	Test Score	Corr Coeff Loss Train	Corr Coeff Loss Test	MAPE Loss Train	MAPE Loss Test	MSE Loss Train	MSE Loss Test
Linear Regression	0.1887	0.0505	0.18871	0.051866	2.0045	2.6437	6.32388	12.519
Randon Forest	0.9573	0.6253	0.97024	0.768843	0.2073	0.8283	1.45085	7.8642
XGB Regressor	0.9822	0.6473	0.98421	0.754504	0.5441	0.9364	0.93621	7.6298
SVM Regressor	-0.005	-0.004	0.00358	0.0044	1.2106	1.6115	7.04017	12.87
MLP Regressor	-59168	-16411	0.00044	0.034211	1040.9	1034.8	1707.82	1645.9

Productivity results for Sugarcane without standarisation								
Model	Train Score	Test Score	Corr Coeff Loss Train	Corr Coeff Loss Test	MAPE Loss Train	MAPE Loss Test	MSE Loss Train	MSE Loss Test
Linear Regression	0.1298	0.0138	0.12983	0.031277	478.71	498.6	2548.56	2831.6
Randon Forest	0.9973	0.988	0.99809	0.988011	8.0277	22.252	140.885	312.51
XGB Regressor	0.9999	0.9869	0.99995	0.989484	11.146	22.135	20.1988	326.53
SVM Regressor with	-0.001	-0.002	0.00041	0.000558	118.97	132.44	2733.81	2853.5
MLP Regressor	-0.218	-0.189	0.00588	0.00814	1169.5	1193.1	3015.46	3108.6

Productivity results for Potato without standarisation								
Model	Train Score	Test Score	Corr Coeff Loss Train	Corr Coeff Loss Test	MAPE Loss Train	MAPE Loss Test	MSE Loss Train	MSE Loss Test
Linear Regression	0.682	0.6466	0.68202	0.654694	2.3366	2.5084	4.43646	4.4096
Randon Forest	0.9466	0.5874	0.95193	0.628544	0.6775	1.902	1.81882	4.7648
XGB Regressor	0.8056	0.386	0.81794	0.471356	2.4388	2.6576	3.46873	5.8121
SVM Regressor	0.2733	0.3023	0.28161	0.314595	4.3648	4.3601	6.70671	6.1957
MLP Regressor	-54541	-62372	0.00648	0.00169	1121.7	1133.4	1837.39	1852.5

Productivity results for Rice without standarisation								
Model	Train Score	Test Score	Corr Coeff Loss Train	Corr Coeff Loss Test	MAPE Loss Train	MAPE Loss Test	MSE Loss Train	MSE Loss Test
Linear Regression	0.1515	0.1878	0.15155	0.427712	0.4009	0.4131	2.39516	0.8483
Randon Forest	0.8357	-6.608	0.89434	0.09579	0.1068	0.3015	1.054	2.5964
XGB Regressor	0.9471	-11.54	0.95773	0.032896	0.4003	0.4605	0.59793	3.3338
SVM Regressor with	0.0188	0.0843	0.02093	0.0977	0.7063	0.6815	2.57569	0.9008
MLP Regressor	-73437	-6E+05	0.0013	2.26E-05	446.22	461.89	704.664	734.25

Production results for Coconut without standarisation								
Model	Train Score	Test Score	Corr Coeff Loss Train	Corr Coeff Loss Test	MAPE Loss Train	MAPE Loss Test	MSE Loss Train	MSE Loss Test
Linear Regression	0.8457	0.7664	0.84571	0.766859	3E+07	3E+07	6.8E+07	8E+07
Randon Forest	0.9855	0.8469	0.98644	0.846998	5E+06	2E+07	2.1E+07	6E+07
XGB Regressor	0.9831	0.8635	0.98387	0.863842	1E+07	2E+07	2.3E+07	6E+07
SVM Regressor with	-0.129	-0.1	0.00201	0.000291	6E+07	5E+07	1.8E+08	2E+08
MLP Regressor	0.1054	0.085	0.12372	0.09656	9E+07	9E+07	1.6E+08	2E+08

Production results for Sugarcane without standarisation								
Model	Train Score	Test Score	Corr Coeff Loss Train	Corr Coeff Loss Test	MAPE Loss Train	MAPE Loss Test	MSE Loss Train	MSE Loss Test
Linear Regression	0.8577	0.8771	0.85769	0.877765	194574	204815	786964	693776
Randon Forest	0.9909	0.9585	0.99158	0.959151	37262	89326	199166	402896
XGB Regressor	0.982	0.9474	0.98277	0.947751	99150	129903	280058	453994
SVM Regressor with	-0.106	-0.123	0.04162	0.057201	717897	735609	2193525	2E+06
MLP Regressor	0.8224	0.8688	0.82411	0.870982	232691	240571	879091	716662

Production results for Rice without standarisation								
Model	Train Score	Test Score	Corr Coeff Loss Train	Corr Coeff Loss Test	MAPE Loss Train	MAPE Loss Test	MSE Loss Train	MSE Loss Test
Linear Regression	0.8983	0.9056	0.89826	0.906061	27372	28022	52590	50949
Randon Forest	0.9947	0.9698	0.99481	0.96981	5166.6	12872	12055.6	28838
XGB Regressor	0.9495	0.9395	0.95009	0.939733	20558	22218	37062.7	40801
SVM Regressor with	-0.188	-0.187	0.03147	0.024952	94396	94379	179677	180660
MLP Regressor	0.8408	0.8388	0.84309	0.841443	37773	38014	65776.8	66578

Production results for Potato without standarisation								
Model	Train Score	Test Score	Corr Coeff Loss Train	Corr Coeff Loss Test	MAPE Loss Train	MAPE Loss Test	MSE Loss Train	MSE Loss Test
Linear Regression	0.9402	0.9265	0.94015	0.926572	10991	12718	30276.5	37614
Randon Forest	0.9936	0.9543	0.99371	0.954806	2748.1	7718.8	9893.53	29669
XGB Regressor	0.9856	0.9483	0.98568	0.948836	6793.3	9541.9	14866.5	31556
SVM Regressor	-0.097	-0.093	0.09606	0.112474	43782	47577	129636	145018
MLP Regressor	0.8167	0.8101	0.85361	0.847494	29711	31814	52985.9	60441

Production results for Wheat without standarisation								
Model	Train Score	Test Score	Corr Coeff Loss Train	Corr Coeff Loss Test	MAPE Loss Train	MAPE Loss Test	MSE Loss Train	MSE Loss Test
Linear Regression	0.9751	0.9769	0.97513	0.977005	22197	23474	38113.9	38756
Randon Forest	0.9973	0.9838	0.99735	0.983962	6282.5	16895	12581.6	32534
XGB Regressor	0.9821	0.9783	0.98224	0.978594	20206	23400	32375.6	37575
SVM Regressor with	-0.177	-0.184	0.07081	0.07053	158214	166861	262166	277733
MLP Regressor	0.9319	0.9323	0.93195	0.932345	40769	42167	63060.5	66409

Production results for Cotton without standarisation								
Model	Train Score	Test Score	Corr Coeff Loss Train	Corr Coeff Loss Test	MAPE Loss Train	MAPE Loss Test	MSE Loss Train	MSE Loss Test
Linear Regression	0.8167	0.7713	0.81673	0.777105	33557	36151	72728.1	73306
Randon Forest	0.9868	0.8952	0.98829	0.896881	6597.6	16652	19519.7	49625
XGB Regressor	0.9618	0.8788	0.96292	0.883001	14554	20229	33208.1	53350
SVM Regressor	-0.144	-0.151	0.04422	0.027333	67883	63184	181675	164446
MLP Regressor	0.678	0.6257	0.68361	0.633408	44763	43999	96405.8	93768