A Suggestive Model For Rice Yield Prediction And Ideal Meteorological Conditions During Crisis

Y. Vijayalata, V.N.Rama Devi, Palakodeti Rohit, G.S.S Raj Kiran

Abstract: The agriculture sector is the backbone of Indian economy. Seventy percent of population in India is dependent on agriculture for their livelihood. With the available resources and infrastructure villagers are able to produce the crop not to the fullest potential yield. In order to augment the efforts of farmers to get higher yields a predictive model was developed. This will help farmers by predicting the yield of the upcoming season by using machine learning and deep learning algorithms such as Deep Neural Network Regressor, Gaussian Process Regressor, Linear Regression and Lasso Regression. This can be done by understanding the relationship between factors that affect crop production (like area, seasonal monthly average temperature, seasonal monthly average humidity, seasonal monthly average rainfall, seasonal monthly average wind speed, seasonal monthly average UV index, seasonal monthly average sun hours, and seasonal monthly average pressure) and crop yield. Prediction will be performed on rice as it is one of the major crops produced in India. If the predicted values are substantially lower than the mean yield then by simulating the weather parameters maximum yield and profit can be obtained.

Keywords: Agriculture, Rice, Prediction algorithms, weather parameters

1 INTRODUCTION

India is the second-largest agricultural land (179.9 million hectares) in the world. Agriculture is one of the main sectors to be impacted by different sources like climatic changes, soil attributes, seasonal changes [1]. India stands number two in the producer and consumer of rice in the world and accounts for 22.3% of global production. Rice contributes to more than 40% of total food grain production and is cultivated throughout the country [5]. Among India's top exported goods basmati rice stands number one. India produces approximately 4.25 million metric tons of basmati rice which is approximately 75% of the total global production.More than half of the basmati rice produced in India is exported.Top Indian basmati rice importers are Iran and Saudi Arabia along with UK and US. Currently, most of India's top rice suppliers and rice exporters are mainly based in regions such as West Bengal, Uttar Pradesh, Andhra Pradesh, Punjab, Tamil Nadu, Orissa, Bihar, and Chhattisgarh. These largest rice producing states account to 72% of the total rice-growing area in India. They hold a share of more than 75% to the total rice production in the country [6]. Crop yield prediction is based on various kinds of data collected and their correlation with yield. This data is fed to the Machine Learning algorithms to train them. Predicting the crop yield can be extremely helpful for farmers. If they have an idea of the amount of yield they can expect, they can make adjustments to their crop prior to harvest, often securing a more competitive price than waiting till the harvest.

- Y. Vijayalata is a Professor in Department of Computer Science, GRIET
- V.N.Rama Devi is an Associate Professor in Department of Mathematics, GRIET
- P. Rohit is final year B.Tech student of Department if Information Technology, GRIET
- G.S.S Raj Kiran is final year B.Tech student of Department if Information Technology, GRIET and corresponding author.
- Mail id is gssrajkiran@gmail.com

The involvement of experts in prediction of crop yield show issues like lack of knowledge about natural events, negation of personal perception and fatigue etc. such issues can overcome by using the models and decision tools for crop yield prediction. Likewise, industry can benefit from yield predictions by better planning the logistics of their business [4].

2. BACKGROUND STUDY

E Manjula et al [2] used different Data Mining techniques and implemented a system to predict crop yield from previous data. This was achieved by applying association rule mining on agriculture data. Data was clustered using k-means clustering algorithm. Brief analysis of crop yield prediction was done. Miss.Snehal S.Dahikar el al [3] considered various situations of climatologically phenomena affecting local weather conditions in various parts of the world. These weather conditions have strong effect on crop yield. Use of Artificial Neural Networks have been illustrated as powerful tools for modelling and prediction, to increase their effectiveness. Usage of Crop prediction methodology was done to predict the suitable crop by sensing various parameter of soil and also parameter related to atmosphere.K. Menaka el al [4] used methods such as Artificial Neural Network, Adaptive Neuro-Fuzzy Inference System, Fuzzy Logic and Multi Linear Regression. These are analysed to know the best methods for crop yield prediction. Various models for crop yield prediction are compared through their parameters such as Root Mean Square Error (RMSE), $R2$, correlation coefficient and Mean Square Error (MSE) to prove Adaptive Neuro-Fuzzy Inference System (ANFIS) prediction model is better than other techniques. Arun Kumar [7] performed descriptive analytics on sugarcane cane crop datasets. Supervised machine learning algorithms were applied to find the actual estimated cost and also a comparative study was done among KNN, SVM and LS-SVM algorithms displaying their accuracy and mean squared error at cross-validation phase.

2 METHODOLOGY

The methodology of the paper consists of the following steps:

3.1 Data Collection

3.1.1 Crop Yield Data

The data set consists of district-wise, season-wise and year wise data on crop covered area and production from the year 2011 to 2015 was obtained from Open Government Data Platform India [8].

3.1.2 Metrological Data

The Metrological data, of district wise, season-wise and year wise data from the year 2011 to 2015 was obtained from worldweatheronline website [9] as on 14th March 2019. The data included behaviour of seasonal monthly average temperature, seasonal monthly average humidity, seasonal monthly average rainfall, seasonal monthly average wind speed, seasonal monthly average UV index, seasonal monthly average sun hours, and seasonal monthly average pressure values.

3.2 Data Preprocessing

3.2.1 Data Normalization

Before applying the algorithms various preprocessing techniques are applied to the data. The data was normalized using Standard Scaler() function from sklearn.preprocessing package. The weather parameters such as humidity, rainfall and temperature, wind speed, UV index, sun hours and pressure were scaled in such a manner that their mean and standard deviation were -1 and 1 respectively.

3.2.2 Algorithm for best parameters selection

3.2.2.1 GaussianProcessRegressor and Lasso Regression

Algorithm model ( )
{Param$_1$ _combinations:= [val$_1$,............ val$_n$];
Param$_2$ _combinations:= [val$_2$,..............val$_n$];.
Param$_n$ _combinations:= [val$_1$,............ val$_n$];
error:=100000
for i$_1$ in param$_1$ _combinations do
for i$_2$ in param$_2$ _combinations do
for i$_n$ in param$_n$_combinations do
model=algorithm(param$_1$=i$_1$, param$_2$= i$_2$,............. param$_n$=i$_n$);
model.fit(x_train,y_train);
if
error>mean_squared_error(model.predict(x_test),y_test):do
error= mean_squared_error(model.predict(x_test),y_test);
print (i$_1$, i$_2$,............., i$_n$);This logic was implemented in order to extract the ideal set of hyper parameters of each algorithm (except Linear Regression) which can extract maximum performance from the model and generate least mean_squared_error. In the above algorithm the param1_combinations, param2_combinations,…..,paramn_combinations are the lists which contain the various values which when taken together generate a combination of hyper parameters with which the machine learning algorithm is trained. In the above algorithm error which is initialized with 100000000 later updated with latest minimum value. When the trained model is generating higher accuracy score then the till now highest acc then we are updating the existing error with new minimum error and printing the combination of parameters which are giving the best results. In this way the best parameters for SVR and Kernel Ridge are chosen.

3 RESULTS AND DISCUSSION

4.1 As a Predictive task

Fig 1 – Fig 5 depict the performance of the respective algorithms by plotting predicted yield values against actual yield values. DNN Regressor and GaussianProcessRegressor (with RBF Kernel) performed the best among all the algorithms.

4.1.1 DeepNeuralNetworkRegressor

- R2 coefficient value is 0.90.
- The MSE (Mean-Squared-Error) is 3070784762.7659636.

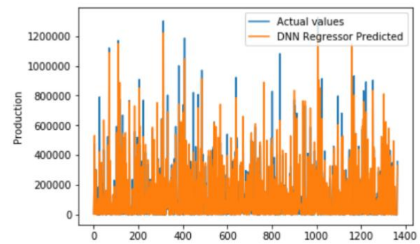

Fig 1

4.1.2 GaussianProcessRegressor

4.1.2.1 GaussianProcessRegressor (with RBF Kernel)

GaussianProcessRegressor(alpha=0.1, copy_X_train=True,kernel=RBF(length_scale=0.1),n_restarts_optimizer=0,normalize_y=False, optimizer='fmin_l_bfgs_b', random_state=None)

- R2 coefficient value is0.90.
- The MSE (Mean-Squared-Error) is 3217811407.7076044.

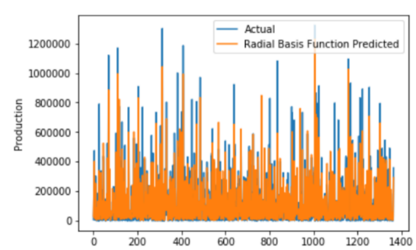

Fig 2

4.1.2.2 GaussianProcessRegressor (with Rational Quadratic Kernel)

GaussianProcessRegressor(alpha=1.0, copy_X_train=True, kernel=RationalQuadratic(alpha=1, length_scale=0.1),n_restarts_optimizer=0, normalize_y=False,optimizer='fmin_l_bfgs_b', random_state=None)

- R2 coefficient value is0.87.

- The MSE (Mean-Squared-Error) is 4049234680.983444.

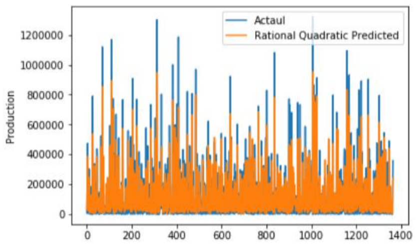

Fig 3

4.1.3 LinearRegressio

LinearRegression(copy_X=True, fit_intercept=True, n_jobs=1, normalize=False)

- R2 coefficient value is 0.73.
- The MSE (Mean-Squared-Error) is 8900231152.876408.

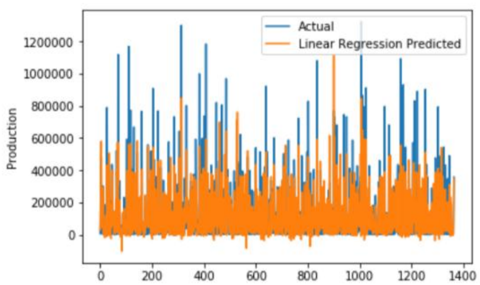

Fig 4

4.1.4 LassoRegression

Lasso(alpha=0.0001,copy_X=True, fit_intercept=True, max_iter=50, normalize=False, positive=False,precompute=False, random_state=None, selection='cyclic', tol=1e-06, warm_start=True)

- It gave an R2 coefficient of 0.73.
- The MSE (Mean-Squared-Error) is 8983018870.928658.

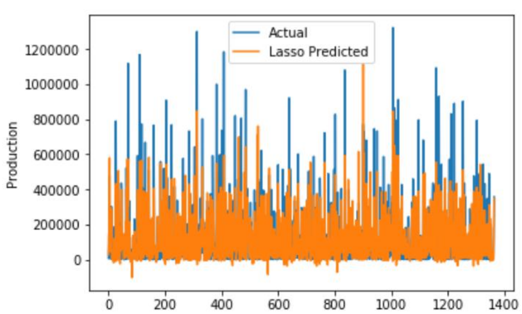

Fig 5

4.2 Suggestions for farmers to maximize crop yield by simulating weather parameters

There will rise certain situations for the farmers where they put their maximum efforts in cultivating the crop but due to bad meteorological conditions the yield is affected, be it heavy rains or a drought or be it inadequate sunlight etc. The proposed model tries to increase the yield in these kinds of situations. Initially the data is fed into the model and the predictions are made. Among the predicted values the anomalies are identified. The anomalies mentioned earlier refer to the scenarios where the predicted value is less the 50% of the mean of the past 4 year's yield of the respective states. These anomalies are taken and considered individually. After considering each anomaly individually the meteorological values under which these anomalies occurred are taken from the dataset. A new dataset is created by manipulating the current meteorological parameter's values. Different combinations with the manipulated values are combined and a new dataset is created. The manipulations are decreasing the parameter's value by 10%, 5% and increasing the value by 5%, 10%. With these manipulated values unique combinations of records (a row consisting the combination of values of meteorological parameters) are created and all together these records make up the dataset. After the creation of the dataset it is fed to the pre-trained neural network and predictions are made. The combination of meteorological parameters for which the prediction of yield is maximum is suggested to the farmer. The changes made to the original value of the parameters by decreasing 10% and 5% and increasing by 5% and 10% because these changes can most probably be accommodated.The proposed model was tested on Karnataka state's rice yield and meteorological data. There were 260(rows) records of Karnataka's data in the dataset. This data was again divided into training and testing data in the ratio 60% and 40% respectively. The model was trained with the training data and the predictions were made to the testing data. Out of the 104 records in the testing data 60 anomalies were identified. Now as mentioned earlier these anomalies parameter values were simulated and then were again given to the model. The highest yield prediction's respective meteorological values were noted down. Table 1 comprises of the actual feature values and the predicted values of the 60 anomalies.Table 2 comprises of the simulated feature values and the respective yield prediction for those feature values for 60 anomalies.

TABLE 1

Area	Hum	Temp	Rainfall	Sun	Values
1495	84.5	22.5	482.48	858	21753.18
1196	76.25	24.75	153.93	851	3616.93
59	75	23.5	276.04	968	5288.73
2063	66.5	25.75	256.71	986	4964.12
361	56.3	23.83	78.25	1591	2553.93
4	62	22.6	137.3	1564	761.74
1558	24	36.3	0.95	984	26697.08
2708	62	25.33	200.13	966	4840.06
612	41.6	29.66	44.91	962	18162.73
73	70	22.5	360.03	1516	6405.47
1802	42.3	28.6	62.27	969	16115.7
1113	76.75	22.75	247.39	944	8474.2
605	47.66	28.3	65.73	952	10343.06
2040	46.66	28.3	15.57	995	13690.62
400	48.33	28	59.04	960	8391.57

963	73.5	23.5	203.69	1020	7363.31
2864	70.75	24.5	96.28	924	8146.37
953	61.6	23.3	214.35	1524	2686.8
644	43.33	29.3	92.28	938	16689.11
682	39.3	30	85	940	21134
3520	49.5	25.33	97.77	1552	7534.9
4617	74.75	23.75	298.85	1034	19083.92
1565	39	30.3	26.69	967	23406.67
3010	59.33	26	180.24	952	6409.93
2126	51.6	28.3	34.28	940	13114.37
1869	33.16	24.83	16.56	1646	7140.07
575	26.3	29.66	87.09	934	16213.72
1410	77	23.25	288.07	990	11900.59
42	62.3	22.6	95.72	1558	672.36
95	72.5	28.6	55.98	879	7129.75
999	39.66	30.3	84.96	924	23169.78
10661	60	26.6	118.06	920	30323.8
7539	81.25	23.5	192.42	974	29461.16
115	49	28.3	27.65	914	8607.52
828	62	23.16	145.75	1568	2313.69
467	53	27.3	131.77	942	4451.61
895	70.75	25	199.08	843	-1769.63
360	51.33	26.6	130.79	959	2804.2
1567	37.33	25.75	19.57	1640	8553.06
7636	73.5	23.75	311.49	1004	24524.2
5054	43.3	30	12.18	966	29585.18
950	46	27.3	53.19	959	7763.8
88	67.5	26.25	185.38	936	45.87
1460	77	23.25	288.07	990	12018.58
7230	76.25	23.75	330.05	1006	26604.02
350	52.3	28	81.57	962	7135.14
6824	68.8	28.3	123.16	1756	14628.08
158	72.26	29.6	70.25	902	9230.02
431	49.33	29	76.54	965	12325
2036	42.16	23.6	129.76	1613	4144.42
6340	72.75	24	236.25	1036	21186.14
3494	65.3	26	107.04	940	10040.67
82	52	27.3	114.09	899	3899.38
96	60.66	27.3	62.91	896	3092.45
1923	53	27.6	107.36	960	9490.49
50	80.25	23	96.07	802	5216.89
97	65	26	138.12	932	-153.67
77	49.33	28	41.77	902	7338.84
5933	64.5	23	170.55	1592	5412.37
344	71.75	24.25	105.13	1064	6275.79

TABLE 2

Area	Hum	Temp	Rainfall	Sun	Values
1345.5	92.95	20.25	530.73	944	35332.86
1076.4	83.88	22.28	138.54	936	34825.56
53.1	82.5	21.15	303.64	871	34935.17
1856.7	73.15	28.32	282.38	888	34382.09
324.9	50.67	21.45	70.42	1591	35009.49
3.6	55.8	20.34	123.57	1486	34847.56
1402.2	21.6	39.93	0.86	885	36243.87
2437.2	55.8	22.8	180.12	870	32417.88
550.8	39.52	32.63	40.42	866	34056.89
65.7	73.5	20.25	396.03	1365	34383.02
1621.8	44.42	31.46	56.04	872	30797.71
1001.7	80.59	20.48	272.13	849	29965.69
544.5	47.66	31.13	59.16	856	29786.36
1836	41.99	31.13	14.01	896	29652.24
360	43.5	30.8	53.14	864	29100.71
866.7	80.85	21.15	183.32	918	28459.45
2577.6	77.82	22.05	86.65	832	28037.36
857.7	55.44	20.97	192.92	1371	29487.98
579.6	39	32.23	83.05	844	26828.91
613.8	35.37	33	76.5	846	27403.63
3168	44.55	22.8	87.99	1397	28727.28
4155.3	74.75	21.38	328.74	931	24187.32
1408.5	39	33.33	24.02	870	25710.95
2709	53.4	28.6	162.22	856	22713.03
1913.4	46.44	31.13	30.85	846	23687.36
1682.1	29.84	22.35	14.9	1481	26497.59
517.5	23.67	32.63	78.38	841	26222.49
1269	84.7	20.92	316.88	891	23199.18
37.8	56.07	20.34	86.15	1402	25121.76
85.5	79.75	31.46	50.38	967	23065.75
899.1	35.69	33.33	76.46	832	24110.4
9594.9	54	29.26	106.25	828	22955.25
6785.1	89.38	21.15	173.18	876	23804.73
103.5	44.1	31.13	24.88	1005	25016.05
745.2	55.8	20.84	131.18	1412	27172.97
420.3	47.7	30.03	118.59	848	23352.92
805.5	77.82	22.5	179.17	927	22661.53
324	46.2	29.26	117.71	863	22877.45
1410.3	33.6	28.32	17.61	1476	26291.22
6872.4	80.85	21.38	342.64	903	19806.69
4548.6	38.97	33	10.96	869	20883.44
855	41.4	30.03	47.87	863	19586.46
79.2	74.25	28.88	166.84	842	18860.12
1314	84.7	20.92	316.88	891	19660.54

6507	83.88	21.38	363.06	905	18444.55
315	47.07	30.8	73.41	865	18365.4
6141.6	75.68	31.13	110.84	1580	21147.49
142.2	79.49	32.56	63.22	992	22567.45
387.9	44.4	31.9	68.89	868	22983.68
1832.4	37.94	21.24	116.78	1452	26200.15
5706	80.03	21.6	259.88	933	20873.21
3144.6	71.83	28.6	96.34	846	20304.44
73.8	46.8	30.03	102.68	989	21191.32
86.4	54.59	30.03	56.62	986	21556.06
1730.7	47.7	30.36	96.62	864	22233.17
45	88.28	20.7	86.46	882	23113.45
87.3	71.5	28.6	124.31	838	20803.89
69.3	44.4	30.8	37.59	993	22009.18
5339.7	70.95	20.7	153.5	1432	23597.32
309.6	78.92	21.82	94.62	958	20165.68

4.3 Preferred seasons in different parts of India

Fig 6 – fig 9 depict the relationship between the area allocated in different seasons and their respective yields.

1. North

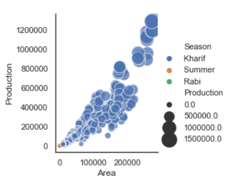

Fig 6

2. South

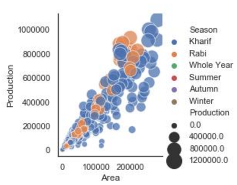

Fig 7

1. East

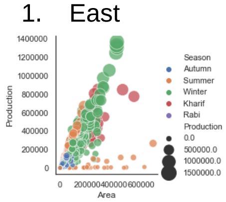

Fig 8

4. West

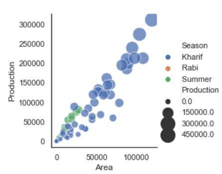

Fig 9

4 CONCLUSION

GaussianProcessRegressor (with RBF kernel) and DNNRegressor with 5 layers and 1000 nodes per layer going through 50 epochs are the algorithms which performed best with R2 coefficients 0.90. The algorithms used in the suggestive model is Gaussian Process Regressor with rbf kernel. The model in most of the situations increased the yield after simulating the parameters. This model would be of aid to the farmer in the situations where the environmental factors are not favourable for cultivation. The changes made in simulation process can be accommodated by artificially creating such environment. The findings in Fig 6 – Fig 9 throw light on how different the approaches of farmers of different parts of India are. Kharif (July – October) season is the most preferred season for cultivating rice and is also the most yield generating season in India.

REFERENCES

[1]. Wu Fan, Chen Chong, Guo Xiaoling, Yu Hua, Wang Juyun. Prediction of crop yield using big data. 8th International Symposium onComputational Intelligence and Design (ISCID).2015;1, 255-260.
[2]. http://www.periyaruniversity.ac.in/ijcii/issue/Vol6No4Mar2017/M5_PID0370.pdf
[3]. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.429.1195&rep=rep1&type=pdf
[4]. http://ijid.informaticspublishing.com/index.php/ijid/article/download/111838/78332
[5]. https://www.researchgate.net/publication/301262798_Production_and_Consumption_of_Grains_in_India
[6]. https://www.bizvibe.com/blog/top-10-largest-rice-producing-states-india-2018/
[7]. https://www.irjet.net/archives/V5/i6/IRJET-V5I6601.pdf
[8]. https://data.gov.in/catalog/district-wise-season-wise-crop-production-statistics
[9]. WorldWeather, https://www.worldweatheronline.com/