

Effectiveness of Machine Learning in Predicting Maize Sowing and Harvesting Time

Srilatha Toomula¹
{srilatha.research@gmail.com¹}

Assistant Professor, Department of Computer Science, RBVRR Women's College, (Autonomous),
Affiliated to Osmania University, Hyderabad, Telangana, India.¹

Abstract. The study aimed to analyse the efficiency of the system made using a machine learning algorithm in predicting the sowing and harvesting time. Further to investigate the opinions of experts and agriculture farmer against the output generated by the system for maize cultivation. The system was designed in such a way to predict sowing time and harvest time using variables: Location, Temperature, Soil Type, Stem Weight, No. of Kernels in an ear, Kernel Weights and Water Content. Results will be displayed in Likert scale whether to sow the maize seeds or not and whether to harvest or not. For this study, opinion-based data was collected from 124 farmers, 113 experts in the agricultural department and compared against 117 data produced by the system. From the analysis, it was understood that there is no significant difference among the opinions for sowing and harvesting of maize. It can be perceived that system opinion matches 97.7% with Farmers Opinion and 96.7% with expert opinion. The estimated R square value is 0.955, meaning the forecasting accuracy of the system is almost 95.5% and the system-generated output has the efficiency equivalent to farmers and experts. For predicting system output using opinion is given by; $\text{System Opinion} = 0.021 + (0.954 \times \text{Farmer Opinion}) + (0.022 \times \text{Expert Opinion})$.

Keywords: ICT in Agriculture, Machine Learning, Experimental & Descriptive Research Design component.

1 Introduction

Agriculture was the catalyst for the development of humanity. India is an agrarian nation with a heavily reliant economy on crop production. As a result, agriculture is the bedrock of India's economy. According to the economic survey of down to earth, part of agriculture in Gross Domestic Product (GDP) is improved to 19.9 per cent in 2020-2021 from 17.8 per cent in 2019-2020 in India and about 50 percent of overall workforce in India are directly and indirectly dependent on agriculture. Predictive approaches for marker-based selection without quantitative trait loci (QTL) mapping will likely be the focus of future applications for complicated traits.[1]

In agriculture, accurate yield estimate and nitrogen control are critical. This research article looks into harvest yield forecasting based on publicly available datasets[2]. This is attributed to machine learning systems' capacity to perform non-linear functions and process a vast number of inputs.

Crop production projections that are accurate and timely are critical for policy choices. In this article, machine learning approaches such as Artificial Neural Networks, decision trees, and Bayesian belief networks are discussed. [3].

Environment, topography, historical, geological, biological, political, and structural, as well as socioeconomic influences, all have an effect on Indian agriculture. Droughts are forecasted and the effects of droughts are mitigated using Support Vector Regression (SVR), Artificial Neural Networks (ANNs), and Wavelet Coupling analysis in this research article [4].

This article highlighted the analysis of production of various crops such as rice, maize, and wheat in various countries, as well as the conclusions taken from satellite data, weather conditions, and risk management [5]. The impacts of pollution, population, solid waste disposal, surface and ground water hydrology, and climate change vary throughout time. The system is set up in such a way that when the data for Location, Temperature, and Soil Type are entered, the system generates a report. Using common land units, the goal of this research was to come up with an accurate, cost-effective, and in-season crop type categorization. [6].

The term "supervised learning" refers to the fact that one teacher is in charge of overseeing the process by which the algorithm is trained using testing models and only used to arrive at an appropriate conclusion with new results. Such forms of supervised learning include artificial neural networks, Bayesian networks, decision trees, help vector machines, ID3, k-nearest neighbour, secret markov model, and so on are all examples of artificial neural networks. Unsupervised machine learning refers to a method in which a large volume of data is supplied to a computer, which then searches for similarities and relationships within them. Unsupervised learning may therefore be used to uncover secret trends in the results. K-nearest neighbour, self-organizing map, partial centred clustering, hierarchical clustering, and k-means clustering are some of the techniques used to cluster data are examples of unsupervised learning algorithms [7].

Machine learning brings together computer science and analytics to increase predictive power. Machine learning methods should be used in agriculture and agricultural production to find correct crop production predictions, since there exists a huge amount of data (ie) produced in agriculture and it is growing every day. Wheat [9] yield is calculated in tonnes per hectare and is classified as total output per unit area. Due to local spatial differences in soil's physical and chemical characteristics, management activities, and regional insect and pathogen damage, within-field yields fluctuate geographically and temporally. Farmers may utilise data on yield variations within a field to improve management choices, benefit, land rental, and insurance value. Yield monitoring, which is the initial step of Precision Agriculture, is one of the most important components for defining management zones (MZ) based on prior yield maps (PA). Yield monitoring assesses the ultimate result of these operations at the end of the growing season, while MZ delineation takes into account soil conditions and previous management practises in the area[8].

2. Review of Literature

Such ML methods are promising and may be employed in computational approaches for maize, such as the more recently popularised linear mixed models, as shown in the use of machine learning (ML) techniques for maize breeding. A new area of ML [1] is emerging among the most cutting-edge technologies for accelerating the introduction of novel genotypes. Many strategic applications of ML in maize breeding, quantitative trait locus

mapping heterotic community assignment, and common genome-wide selections are being addressed in the literature. Corn is one of the world's most valued grains, along with rice and wheat, and a major source of calories for humans. Development of genotypes suited to deteriorating environmental circumstances, primarily drought, that must be cultivated in marginal law and changing climatic conditions for agricultural production[9].

In order to devise crop management systems, researchers used advanced machine learning and modelling tools, as well as studied different methods of a relaxation platform for winter wheat crop management.

Root Mean Square Error (RMSE), Root Relative Square Error (RRSE), Normalized Mean Absolute Error (NMAE), and Correlation Factor (R) were the four accuracy measures employed in this paper to assess the yield prediction model [10].

By taking into account a variety of independent variables like temperature, soil, water level., a model is constructed to forecast Wheat production before harvest in Punjab's seven districts and analyzed the results [11]. This research study focused on a field experiment with data sets to compare the efficacy of HYDRUS-2D with different machine-learning models[12]. Within-field zone yield varies spatially and owing to regional pest and pathogen damage, management actions, and local spatial differences in soil physical and chemical characteristics. Farmers may utilise data on yield variations within fields to improve their management, revenue, land leasing, and insurance value [15]

Yield monitoring, which is one of the most essential components for establishing management zones (MZ) utilising prior yield charts, is the initial step in PA. High-resolution satellite imagery data is used to predict corn crop yields[13]. In large fields with substantial geographic variability, traditional agricultural yield monitoring methods such as quadratic frame sampling are time-consuming, disruptive, labor-intensive, and incorrect. Crop yield monitoring devices mounted on harvesters are now available and can provide yield charts, but only at the end of the season. As a consequence of the exponential development of Machine Learning and the requirement for agricultural production monitoring and prediction, many academics are utilising satellite and aerial Machine Learning data to analyse within-field variability. This article discusses the applications of the ANN model, which solves hydrological issues and provides reliable findings[14].

Using archival Sentinel-2 data, the Green Normalized Difference Vegetation Index (GNDI) at crop stage is utilised to assess within-field variability of maize production from previous seasons[15]. Using remote sensing pictures and machine learning classifiers, identification large-scale crop distribution is focused in this paper [16]

3. Methodology

The system was designed in such a way to predict sowing time and harvest time using variables such as; Location, Temperature, Soil Type, Stem Weight, No. of Kernels in an ear, Kernel Weights and Water Content in Kernel. The system function in such a way that, when the Location, Temperature and Soil Type data is given as an input it provides results in Likert scale (Highly Not Recommended – Highly Recommended) whether to sow the maize seeds or not. Similarly, when the sample data of the Stem Weight, No. of Kernels in an ear, Kernel Weights and Water Content in Kernel is given it provides a result in Likert whether to harvest or not. For the study, opinion-based data was collected from 124 farmers, 113 experts in the agricultural department and compared data against 117 data produced by the system.

The study was conducted to analyse the efficiency of the system made using a machine learning algorithm in predicting the sowing and harvesting time. Further to investigate the opinions of experts and agriculture farmer against the output generated by the system for maize cultivation.

4. Analysis and Interpretation

The analysis was carried to identify, whether there is any significant difference in the opinion's among the experts, farmers and system-generated output at any point in time.

From the Table No.1, the estimated significance value is 0.668 which is greater than 0.05 [Std. Value] meaning the null hypothesis is accepted. Therefore, there is no significant difference in opinion among the experts, farmers and system-generated output at any point of time for sowing and harvesting of maize. The graph further indicates that the opinion is symmetrical and normalized.

Having found the opinion between experts, farmers and system-generated output at any point of time are the same. The analysis was carried out using the Pearson correlation test to identify the level of relationship between experts, farmers with system generated output.

From Table No. 2, the estimated significance value is less than 0.05 [Std. value] Meaning the null hypothesis is rejected. Therefore, there is a significant relationship between the opinions made by expert, farmer and system. From the correlation value, it can be interpreted that there can be a strong positive relationship between the opinions made by expert, farmer and system. Further, it can be perceived that system opinion matches 97.7% with Farmers Opinion and 96.7% with expert opinion on sowing and harvesting of maize.

Table No. 1: ANOVA Test on opinions among the experts, farmers and system-generated output

ANOVA					
Opinion					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.697	2	.848	.4	.668
Within Groups	736.555	351	2.098	04	
Total	738.251	353			
Correspondence Table					
Opinion	Farmer / Experts / System				
	Farmers	Experts	System	Active Margin	
Highly Not Recommended	27	27	28	82	
Not Recommended	26	27	28	81	
Neither Recommended nor Not Recommended	20	17	15	52	

Recommended	25	24	25	74
Highly Recommended	26	18	21	65
Active Margin	124	113	117	354

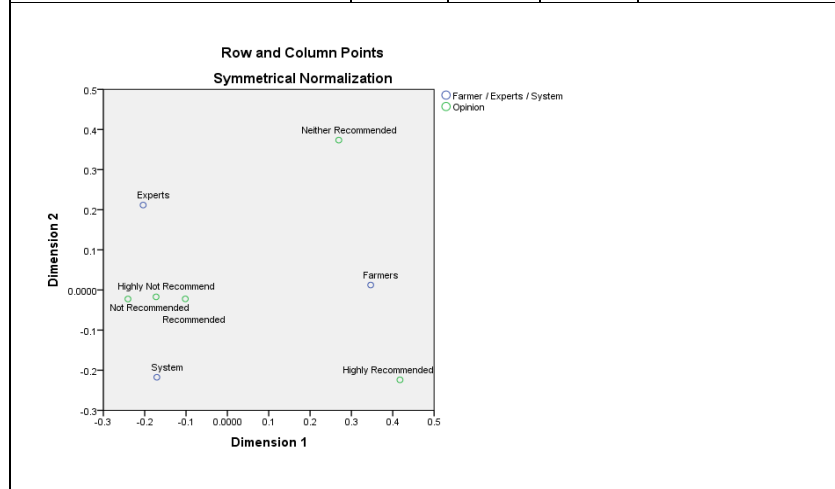


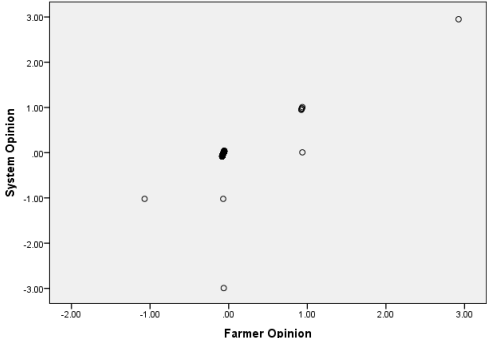
Table No. 2: Correlation – Level of Relationship between the opinion of experts, farmers and system

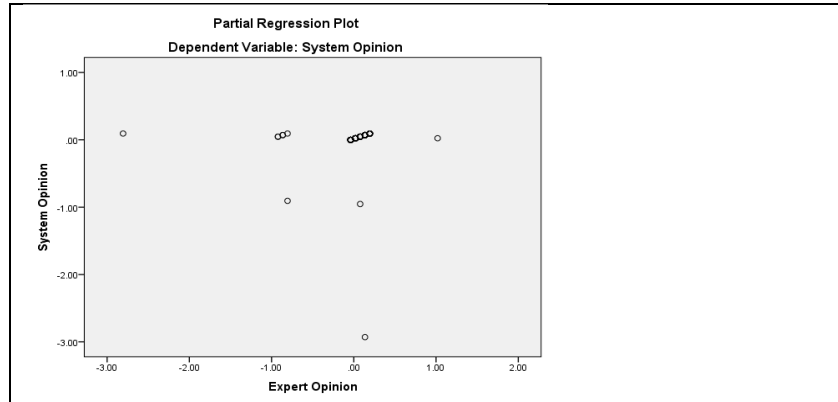
Correlations		Farmer Opinion	Expert Opinion	System Opinion
Farmer Opinion	Pearson Correlation	1	.967**	.977**
	Sig. (2-tailed)		.000	.000
	N	113	113	113
Expert Opinion	Pearson Correlation	.967**	1	.947**
	Sig. (2-tailed)	.000		.000
	N	113	113	113
System Opinion	Pearson Correlation	.977**	.947**	1
	Sig. (2-tailed)	.000	.000	
	N	113	113	113

** . Correlation is significant at the 0.01 level (2-tailed).

*Source: (Primary data)

Table No.3: Regression – Efficiency of the system

Model Summary						
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate		
1	.977a	.955	.955	.31104		
a. Predictors: (Constant), Expert Opinion, Farmer Opinion						
b. Dependent Variable: System Opinion						
ANOVA						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	228.491	2	114.245	1180.899	.000b
	Residual	10.642	110	.097		
	Total	239.133	112			
a. Dependent Variable: System Opinion						
b. Predictors: (Constant), Expert Opinion, Farmer Opinion						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.021	.065		.318	.751
	Farmer Opinion	.956	.079	.957	12.071	.000
	Expert Opinion	.022	.081	.021	.271	.000
a. Dependent Variable: System Opinion						
<p style="text-align: center;">Partial Regression Plot</p> <p style="text-align: center;">Dependent Variable: System Opinion</p>  <p style="text-align: center;">Farmer Opinion</p>						



*Source: (Primary data)

The estimated R square value is 0.955, meaning the forecasting accuracy of the model is almost 95.5%. The ANOVA Significance is less than 0.05 [Std. Value] Meaning the model is fit. Further, the coefficients significance value is less than 0.05. This indicates that the system generated output can be predicted with expert and farmer opinion i.e. the system-generated output has the efficiency equivalent to experienced farmers and experts. The graphical interpretation using the scatter plot method on the opinion illustrates that most of the opinion made by experts and farmers are similar to the output generated by the system

The regression equation for predicting system output using expert and farmer opinion is given by:

$$\text{System Opinion} = 0.021 + (0.954 \times \text{Farmer Opinion}) + (0.022 \times \text{Expert Opinion})$$

In agriculture, a growing number of machine learning implementations are required, for which a large amount of data presently available from many sources may be analysed to reveal hidden information. This will be a well-researched area with the potential to grow in the future. Agricultural crop forecasting is aided by the integration of computer science and agriculture. Preharvest crop forecasts must be based on objective methods. Over traditional forecasting techniques, developing an appropriate model offers many benefits. More information on the analysis may be found in Table 1.

5. Conclusion

From the analysis point of view, there is no significant difference in opinions collected from the experts, farmers and system-generated output at any point of time for sowing and harvesting of maize. Further, it can be perceived that system opinion matches 97.7% with Farmers Opinion and 96.7% with expert opinion in sowing and harvesting of maize. The estimated R square value is 0.955, (i.e.) the forecasting accuracy of the system is almost 95.5%.

References

- [1] R. Bernardo, "Molecular markers and selection for complex traits in plants: Learning from the last 20 years," *Crop Sci.*, vol. 48, no. 5, pp. 1649–1664, 2008, doi: 10.2135/cropsci2008.03.0131.
- [2] S. Lekashri and R. Manikandan, "Efficient Crop Yield Recommendation System Using Machine Learning For Digital Farming," vol. 10, no. 1, pp. 906–914, 2021.
- [3] S. Mishra, D. Mishra, and G. H. Santra, "Applications of machine learning techniques in agricultural crop production: A review paper," *Indian J. Sci. Technol.*, vol. 9, no. 38, 2016, doi: 10.17485/ijst/2016/v9i38/95032.
- [4] A. Belayneh and J. Adamowski, "Drought forecasting using new machine learning methods," *J. Water L. Dev.*, vol. 18, no. 9, pp. 3–12, 2013, doi: 10.2478/jwld-2013.
- [5] E. Biffis and E. Chavez, "Satellite Data and Machine Learning for Weather Risk Management and Food Security," *Risk Anal.*, vol. 37, no. 8, pp. 1508–1521, 2017, doi: 10.1111/risa.12847.
- [6] Y. Cai et al., "A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach," *Remote Sens. Environ.*, vol. 210, no. February, pp. 35–47, 2018, doi: 10.1016/j.rse.2018.02.045.
- [7] [Mohamed Alloghani](#), Dhiya Al-Jumeily Obe, [Jamila Mustafina](#), [Abir Hussain](#), Ahmed J. Aljaaf, "A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science", © Springer Nature Switzerland, ISBN 978-3-030-22475-2, January 2020, DOI: [10.1007/978-3-030-22475-2_1](https://doi.org/10.1007/978-3-030-22475-2_1)
- [8] R. Kaundal, A. A. Kapoor, and G. P. S. Raghava, "Machine learning techniques in disease forecasting: A case study on rice blast prediction," *BMC Bioinformatics*, vol. 7, pp. 1–16, 2006, doi: 10.1186/1471-2105-7-485.
- [9] X. E. Pantazi, D. Moshou, T. Alexandridis, R. L. Whetton, and A. M. Mouazen, "Wheat yield prediction using machine learning and advanced sensing techniques," *Comput. Electron. Agric.*, vol. 121, pp. 57–65, 2016, doi: 10.1016/j.compag.2015.11.018.
- [10] A. Gonzalez-Sanchez, J. Frausto-Solis, and W. Ojeda-Bustamante, "Predictive ability of machine learning methods for massive crop yield prediction," *Spanish J. Agric. Res.*, vol. 12, no. 2, pp. 313–328, 2014, doi: 10.5424/sjar/2014122-4439.
- [11] U. Saeed, J. Dempewolf, I. Becker-Reshef, A. Khan, A. Ahmad, and S. A. Wajid, "Forecasting wheat yield from weather data and modis ndvi using random forests for punjab province, Pakistan," *Int. J. Remote Sens.*, vol. 38, no. 17, pp. 4831–4854, 2017, doi: 10.1080/01431161.2017.1323282.
- [12] F. Karandish and J. Šimůnek, "A comparison of numerical and machine-learning modeling of soil water content with limited input data," *J. Hydrol.*, vol. 543, pp. 892–909, 2016, doi: 10.1016/j.jhydrol.2016.11.007.
- [13] N. R. Peralta, Y. Assefa, J. Du, C. J. Barden, and I. A. Ciampitti, "Mid-season high-resolution satellite imagery for forecasting site-specific corn yield," *Remote Sens.*, vol. 8, no. 10, 2016, doi: 10.3390/rs8100848.
- [14] Rakesh Tanty and Tanweer S. Desmukh, "Application of Artificial Neural Network in Hydrology-A Review," *Int. J. Eng. Res.*, vol. V4, no. 06, pp. 1–9, 2015, doi: 10.17577/ijertv4is060247.
- [15] A. Kayad, M. Sozzi, S. Gatto, F. Marinello, and F. Pirotti, "Monitoring within-field variability of corn yield using sentinel-2 and machine learning techniques," *Remote Sens.*, vol. 11, no. 23, 2019, doi: 10.3390/rs11232873.
- [16] N. Yang et al., "Large-scale crop mapping based on machine learning and parallel computation with grids," *Remote Sens.*, vol. 11, no. 12, pp. 1–22, 2019, doi: 10.3390/rs11121500.
- [17] .