Data-Driven Farming

SAMRIDH GUPTA[1], RAUNAK JASRASARIA[2]

Abstract- Out of all the three sectors of the Indian Economy, the primary sector has not enjoyed the benefits of technological advancements in recent years as much as the secondary and tertiary sectors have. Unfortunately, the agricultural sector, on which more than 70% of Indian rural households depend, has been left out of this revolution. Many programs and initiatives have been launched by governing bodies to educate farmers and provide them with technical aid to maximize their harvest. However, not much emphasis has been laid on matching the supply of various crops to their respective market demands. The lack of any such policy has resulted in a surplus supply of crops leading to the wastage of food and farmers' money. Our system will guide farmers about how much crop they should produce in a particular year so that there is minimum or now a stage of crops.

Indexed Terms- Support Vector Machine (SVM), crop production, government data for crop production

I. INTRODUCTION

One of the most significant professions in India is agriculture. The nation's whole economy depends on this sector because it is the most diverse. In order to feed the nation's 1.3billion inhabitants, more than 60% of the land is used for agriculture. Implementing cutting-edge agricultural technologies is therefore essential since it will benefit our farmers.

Machine learning has been used in the classification in many areas namely; agriculture, defect prediction, text classification. The prediction of crop production is one of the emerging topics, and numerous models have been developed and shown to be successful [1].Agricultural production depends on variety of features namely; climate, soil, seed, moisture, weather and use of fertilizers. This shows that predicting crop yields is a difficult undertaking requiring several intrica test ages. The actual crop yield is predicted with the help of machine learning models.

The most significant of machine learning's applications are being developed with the help of a rapidly evolving approach that aids all industries in making informed decisions. Most contemporary systems gain from model evaluation prior to deployment.

Since there were more parameters in the training period than in the actual training period, this is another factor that affects the prediction.

From the collection of historical data, this research offers a precise approach that aids in crop forecast. The information is provided using past data on various crop yields and demand. We created an application that runs the algorithm and displays a list of crops along with an estimate of their yield value.

II. LITERATURE REVIEW

Several studies have been done using machine learning and various algorithms in the field of agriculture to assist farmers in crop production.

In study [1] states that one method of estimating crop yields is to compute crop yields based on previously gathered historical data, such as precipitation, soil, and prior crop yields. Farmers will be able to anticipate crop production with the use of this prediction. Utilizing three data sets—clay, precipitation, and production data from the state of Karnataka—an integrated data set is constructed and three machine learning algorithms are applied order to determine the classification Accuracy. It is implemented using Python programming and the spyder tool with machine learning algorithms namely; K-Nearest Neighbour (KNN), Support Vector Machine (SVM), and Decision tree algorithms.

N. Manjunathanetal. [3] built a machine learning model that can precisely estimate rice crop

production is the major goal of their proposal. For this, they have employed the Support Vector Machine (SVM) method. This proposal's primary goal is to create a machine-learning model that can precisely estimate rice crop production.

Authors in [4] designed a method where she utilizes the support vector machine algorithm and k-nearest neighbors to forecast the best crop. Her methodology aids in determining the most practical crops to farm for a certain place in light of the existing environmental circumstances. The data set utilized for this research was compiled from a number of reliable internet sources, including Kaggle.com, data.govt.in, and many others.

III. METHODOLOGY

In an aim to benefit the farmers, our main motto is to create a machine learning model which will try to predict the market demands of a crop keeping in mind various features such as the previous Year's demand. We will be using data from the government records like the growing population trend, production of various crops, and their changing demand over the years. This will be used to find the surplus and ultimately predict the demand for future years. The data we will take into consideration will be for rice, maize, and wheat as they are primary growths for the country. The data collected will be divided into training and testing tests which will help us train the model and then find its accuracy.

- The data set that we are using is shown below: (Records are from1960-2020)

The data use discollected from various official government websites and then arranged in a tabular manner, which when converted to a CSV format isused by the Machine learning model.

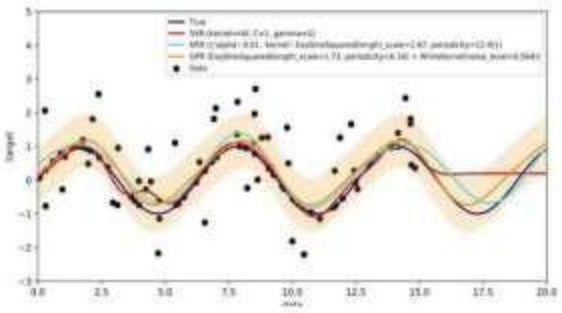
The data contains the population trend of the country over the years and details like the production and demand of various major crops in India. Using this data and a few other factors the machine learning model calculates the demand for the upcoming years.

year	population	wheat_pri	rice_prod	maize_pr	wheat_de	rice_dem	maize_dem
2020	1.38E+09	107.59	121.46	28.2	99.5	104	28.2
2019	1.366E+09	103.6	118.43	28	96.11	102.34	28
2018	1.353E+09	99.87	116.48	28.5	95.63	99.16	28.5
2017	1.339E+09	98.51	112.76	26.7	95.68	98.67	26.7
2016	1.325E+09	92.29	109.7	24.9	97.23	95.84	24.9
2015	1.31E+09	86.53	104.41	23.55	88.55	93.45	23.55
2014	1.296E+09	95.85	105.48	22.35	93.1	98.24	22.35
2013	1.281E+09	93.51	106.65	19.6	93.85	98.73	19.6
2012	1.266E+09	94.88	105.24	17.5	83.82	93.97	17.5
2011	1.25E+09	86.87	105.3	17.2	81.41	93.33	17.2
2010	1.234E+09	80.8	95.98	18.1	81.76	90.2	18.1
2009	1.218E+09	80.68	89.09	15.1	78.15	85.5	15.1
2008	1.201E+09	78.57	99.18	17	70.92	91.08	17
2007	1.183E+09	75.81	96.69	14.2	76.42	90.46	14.2
						96.7	13.9

- Machine Learning Algorithm for prediction: The outputs of machine learning prediction algorithms are based on pre-entered data and have highly optimized estimates. Predictive analytics calculates the likelihood of future events based on past data using data, statistical algorithms, and machine learning techniques. By looking beyond only what has already happened, the goal is to give the best accurate prediction of what will happen in the future.

In our System, we have used Support Vector Machines to predict the yield of different crops.

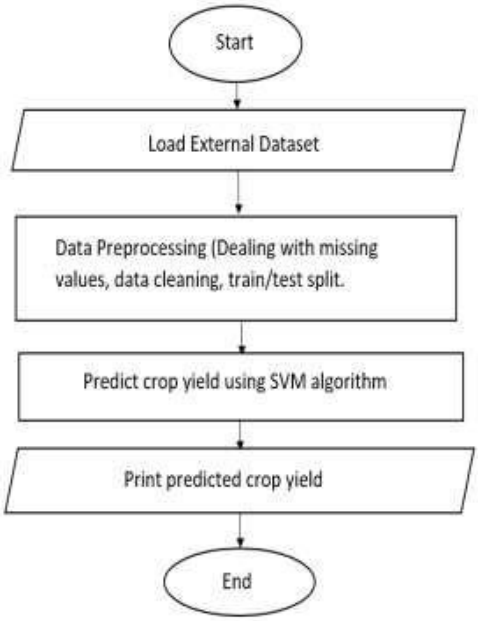
- Support Vector Machine Using supervised a machine learning algorithm that is (SVMs) classification or regression tasks are possible. A high-dimensional space's hyper planet hat maximum separates the classes is what an SVM seeks to locate. In the case of regression, an SVM, like in the classification scenario, finds a hyper plane that maximum separates the data points. However, the objective is to produce a smooth curve that can be used to generate predictions for new data points rather than categorizing the points into separate groups.



Support Vector Regression (SVR), on the other hand, is a variation of the SVM technique that is applied to regression tasks. Finding a smooth curve that can be used to predict fresh data points is the aim of SVR. Instead of maximizing the separation between the classes as in the classification instance, SVR aims to reduce the difference in values between anticipated and actual values.

Support vector regression is a robust and adaptable machine learning approach that is suitable for regression problems, especially when working with noisy and non-linear data. It is frequently used in many fields, including finance, engineering, and medicine. Support Vectors are the data points or a vector that lies closest to the hyper plane. They have high impact on hyper plane.

- The flow chart for crop production is shown below:



IV. RESULT

In order to analyze the performance a model, there are various metrics available. We have chosen ther2 score metric for the accuracy prediction. It is a measure of the goodness of fit of aregression model and provides a score of how well the model is able to predict the dependent variable based on the independent variable(s).

The R2 score goes from 0 to 1, with a value of 0 indicating that the model is completely incapable of forecasting the dependent variable and a score of 1 indicating that it can do so with absolute certainty. A score between 0 and 1 indicates that the model is able to predict the dependent variable to some degree, with higher scores indicating better predictions.

The formula for calculating the R ^ 2 score is: R^2 = 1-(Sresidual /Stotal)
Where Total is the total sum of squares of the differences between the observed values and the mean of all observed values, and Sresidual is the sum of squares of the residuals (the discrepancies between observed values and the predicted values).

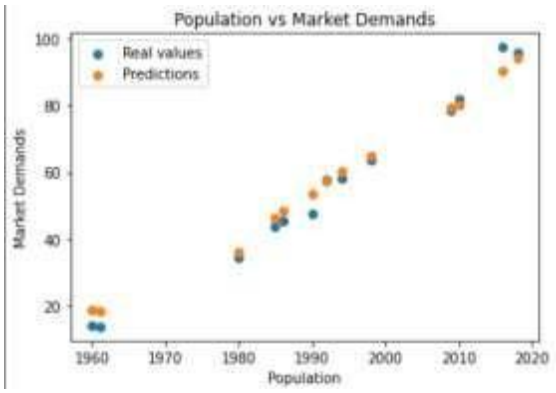
In this study, the above-mentioned support vectormodelwascreatedtofindthequantityofcropsrequiredforfutureconsumption.Themodelwascreated with an accuracy of 97%.

CONCLUSION

In this study, crop forecasting and yield estimation are done using machine learning approaches. The crop yield was calculated and predicted using the SVM machine learning technique. Created a method to anticipate crops based on the gathering of historical data, with an accuracy of about 97%. The suggested method aids farmers in determining which crop to plant in the field and how much of it. This work is done in an effort to learn more about the crops that may be used for effective and helpful harvesting. Farmers all around India would benefit from precise forecasts of several specific crops in various districts. As a result, crop yield rates are increased, which is good for the Indian economy.

Currently, our farmers are not utilizing technology

and analysis effectively, therefore there is a chance that they may select the wrong crops to plant, which would reduce their income. In order to predict which crop would be the best fit for a particular piece of land and to provide pertinent information about the nutrients that must be added, the seeds that must be used for cultivation, the anticipated yield, and the market price, we developed a farmer-friendly system with a graphical user interface (GUI).This inspires farmers to make good decisions as a result.



V. FUTURE SCOPE

Machine learning has a lot of potential in agriculture for crop prediction and crop production enhancement. Crop yield prediction, insect identification, irrigation optimization, etc. are a few potential uses of machine learning in agriculture.

In the coming years, we can extend this algorithm to include more crops so as to make it more useable and practical in nature. Moreover, we can include various other features into consideration as well, such as soil, weather, rainfall, etc. We can also make a login page for different farmers to login and make their accounts. We can also attempt to put data-independent systems into place so that, regardless of the format, our system should function accurately.
In general, the system will be able to meet farmers' needs in the near future.

REFERENCES

[1] Renuka, S. T. (2019). Evaluation of machine learning algorithms for crop yield prediction. *Int. J. Eng. Adv. Technol.(IJEAT)*, *8*(6).

[2] Kamath, P., Patil, P., Shrilatha, S., & Sowmya, S. (2021). Crop yield forecasting using data mining. *Global Transitions Proceedings*, *2*(2), 402-407.

[3] Manjunathan, N., Rajesh, P., Thangadurai, E., & Suresh, A. (2020). Crop yield prediction using linear support vector machine. *Eur J Mol Clin Med*, *7*(6), 1-7.

[4] Van Klompenburg, T., Kassahun, A., &Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, *177*, 105709.

[5] Hina, F., & Hasan, M. T. Agriculture Crop Yield Prediction Using Machine Learning.

[6] Yost, M. A., Kitchen, N. R., Sudduth, K. A., Sadler, E. J., Drummond, S. T., & Volkmann, M. R. (2017). Long-term impact of a precision agriculture system on grain crop production. *Precision agriculture*, *18*(5), 823-842.

[7] Mishra, S., Mishra, D., & Santra, G. H. (2016). Applications of machine learning techniques in agricultural crop production: a review paper. *Indian J. Sci. Technol*, *9*(38), 1-14.

[8] Braunack, M. V. (2013). Cotton farming systems in Australia: factors contributing to changed yield and fibre quality. *Crop and Pasture Science*, *64*(8), 834-844.

[9] Papageorgiou, E. I., Markinos, A. T., &Gemtos, T. A. (2011). Fuzzy cognitive map based approach for predicting yield in cotton crop production as a basis for decision support system in precision agriculture application. *Applied Soft Computing*, *11*(4), 3643-3657.

[10] "SupportVectorRegressionInMachineLearning" https://www.analyticsvidhya.com/blog/2020/03/support-vector-regression-tutorial-for-machine-learning/

[11] Rao, M. S., Singh, A., Reddy, N. S., & Acharya, D. U. (2022). Crop prediction using machine learning. In *Journal of Physics: Conference Series* (Vol. 2161, No. 1, p. 012033). IOP Publishing.https://flask.palletsprojects.com/en/2.2.x/

[12] https://www.data.gov.in

[13] Ashapure, A., Jung, J., Chang, A., Oh, S., Yeom, J., Maeda, M., ... & Smith, W. (2020).

Developing a machine learning based cotton yield estimation framework using multi-temporal UAS data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *169*, 180-194.

[14] Venugopal, A., Aparna, S., Mani, J., Mathew, R., & Williams, V. (2021). Crop Yield Prediction using Machine Learning Algorithms. *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NCREIS*, *9*(13).