

GSJ: Volume 9, Issue 8, August 2021, Online: ISSN 2320-9186 www.globalscientificjournal.com

AI for sustainability:-Spice Yield Prediction using Indian Climatic Parameters.

Dr Nupur Giri HOD,Dept.Of Computer Engineering VESIT Mumbai, India <u>nupur.giri@ves.ac.in</u> Mr Siddarth Bugtani Dept. Of Computer Engineering VESIT Mumbai, India 2018.siddarth.bugtani@v es.ac.in Ms Yashvi Hiranandani Dept. Of Computer Engineering VESIT Mumbai, India 2018.yashvi.hiranandani @ves.ac.in

Mr Vishwesh Jagtap Dept. Of Computer Engineering VESIT Mumbai, India 2018.vishwesh.jagtap@ve <u>s.ac.in</u>

Abstract--India is a land of spices and their successful husbandry depends upon the area provided for cultivation and the climatic conditions. India is one of the chief exporters of spices. Therefore yield forecasting plays a vital role in export planning and policy decisions.Over the years research and development in machine learning has provided a significant contribution to industries. This project aims to create a model to predict the yield of spices based on the region of cultivation and various other input parameters which would help the farmers to foresee the yield and take appropriate measures for storage and plantation of these crops. We have also provided the prices forecast of the spices to determine whether it is worth investing in its plantation for the current year or not. Algorithms such as **Random Forest Regressor, Stochastic Gradient** Descent and KNN were applied to the dataset after performing preprocessing steps. Advanced regression techniques like Kernel Ridge, Lasso and ENet algorithms have been used to predict the yield along with Stacking Regression which

Mr Kabir Rajpal Dept. Of Computer Engineering VESIT Mumbai, India 2018.kabir.rajpal@ves.ac. <u>in</u>

gives us better results in yield prediction. The results have been expressed using the mean absolute error and the r2 score.



Fig 1. Export % of each spice in India

Keywords:-Kernel Ridge Regressor, Wholesale Price Index(WPI), Climate, Soil moisture.

I. Introduction:-

Farming is the spine for each nation's economy primarily like India, which has consistently rising interest for nourishment because of rising populace,

GSJ© 2021 www.globalscientificjournal.com and in house demand. Agriculture is not only the primary occupation of our nation, but has also been the backbone for the economy since ages. Production and cultivation play an important role in it. At times the productivity of crops is sufficient but the export is lagging due to several conditions. Thus the demand for food keeps on increasing. Spices constitute a major place in the kitchen as they bring out the unique natural taste of cuisines and could be used to change the look of food to make it more attractive and pleasing to the eyes .Along with its various advantages, spices have several health benefits. Since years, spices add up to the taste and make the home-cooked food taste more better, no only in india but all around the world. Spices have a lot of nutritional value thus providing some major health benefits. The Economic times [1] pointed out that, "During this covid-19 pandemic, out of India's total crop production, 90% produce was of spices with the total export worth 250 billion rupees in one year." As per the financial express[2] The cost of export of various spices increased in the first six months of 2021 as compared to the past one year. Though In value terms, as stated in The Business Today[3], "India exported spices worth Rs 15,882." Though there is a 8 percent growth as compared to last year, the production and export difference is still notable. The main reason for this difference is the lack of knowledge. Mainly farmers don't know where the produce will be more, they have been practising traditional things for over years now. Researchers are trying their best to bridge this gap. Hence in order to obtain finest results in predicting the crop yield, some of the machine learning algorithms can be implemented. Machine learning is basically the branch of computer engineering where the machine is trained to work as a human brain and thus give great results, which in turn maximizes the profits and increases the success. With machine learning, the computer is made to do the things that the human does. Crop production is a complicated development that's influenced by soil and environmental condition input parameters and let alone the traditional techniques don't seem to be helping enough so with the help of these machine learning methodologies the task can be done greatly. Though agriculture input parameters vary from field to field and farmer to farmer, proper planning, learning, communication, reasoning, knowledge and perception have made this

possible. Modern technological advancement in the field of yield prediction also helps the farmers with future cost prediction. While predicting yield of crops, namely spices, we do classification along with prediction. This prediction could be price prediction or prediction of crops suitable in a particular area. In this paper we have taken 9 spices grown all around India and predicted their yield and future prices from previous available data as well as various environmental data. The 9 spices which we chose on the basis of productivity and usage are as follows Black pepper, cumin, rapeseed and mustard, coriander, turmeric, ginger, garlic, dry chilli and cardamom.Black pepper since India is one of its major producers. Ginger and garlic hold a huge position in the urban market. The reason behind choosing Cardamom is that as per FBNews.com[4] it is the third most expensive spice when seen globally, whereas turmeric is the commonly used Indian spice. The Mordor Intelligence[5] states that, "19% of the rural market is dominated by dry chillies." Coriander being the oldest spice and cumin being dense in iron, helps in iron deficiency. With our system, the farmers will be handed in with all the necessary information specific to a region. Right from the best crop for that region to the future prices. Thus from this yield prediction, we provide the necessary information to the farmers thus helping increase their yield and also letting them know about the types of crops to be cultivated during that climate.

II. Related Work:-

- SML Venkata Narasimhamurthy & et al [6]. "Rice Crop Yield Forecasting Using Random Forest Algorithm". In this paper, this algorithm is used to predict the yield of rice, by considering the climatic conditions as the parameters thus helping farmers in making decision. Climatic conditions like rainfall, temperature and data like perception and rice production in tonnes have been taken into consideration. Maximum accuracy of 85. 89% is obtained by using this method.
- P. Priya*1 & et al [7]. "Predicting the Crop Yield Using Machine Learning Algorithm". In this paper, a Random forest algorithm has been used to check the yield of the crop as

per the hectares, before farming onto the field. Data sets such as rainfall, perception, temperature, production is taken into account. Decision tree classifier has been used. Here, accuracy can be predicted by relating the resultant class value with the test data set.

- (Filippi et al., 2019)[8] proposed random forest models combined with STC to predict wheat, barley, canola yield for three different seasons. LOFOCV and LOFYOC cross validation methods are used to measure the prediction quality of the models. The RMSE of the models suggest that more data driven models can be developed using machine learning algorithms to predict crop yield.
- (Tamsekar et al., 2019)[9] proposed a crop selection prediction model using GIS data, irrigation data, chemical data, soil data and yield data. Machine learning models such as CART, KNN, RF and SVM are used to build

III. Materials and Methods:-

A. Datasets:-

- Crop Data: Obtained from data.gov.in.The columns in the dataset are State,District,Crop,Year, Season,Area and Production.
- 2. *Climatic Data:* The dataset contains values of temperature and humidity of states from 1989 to 2015.
- Soil Data: This dataset contains soil moisture and soil ph level values of the States.
- 4. *Rainfall Data*: Rainfall data was collected from Indian Water Portal. Consists of data for every month from 2000 to 2018.

the prediction model. PCA is used to improve the model through dimension reduction. The SVM with PCA showed better results with an accuracy of 77%.

- 5. Saranya C P, Guru Murthy B [10] "A SURVEY ON CROP YIELD PREDICTION USING MACHINE LEARNING ALGORITHMS". This paper focuses on a concise relative work of various papers that deals with different techniques used to evaluate the crop yield. It aims at predicting crop yield (Wheat) by considering and examining the datasets of previous years of the crop. In addition, this paper presents various existing techniques to audit crop yield. It also constitutes the contrast of various algorithms along with its benefits and drawbacks. The algorithms discussed are Random Forest, Linear Support Vector Machine, Support Vector Machine, k-Nearest Neighbour, Artificial Neural Networks, Decision Tree, Multiple Linear Regression.
- 5. *WPI values*: Obtained from data.gov.in,this dataset has WPI values for most of the crops including a few spices.

B. Data Mapping:-

The project consists of 2 sections.For the first section Crop, Climate, Rainfall and Soil data were merged and the spices were segregated from the merged dataset.

For the second part,WPI values of spices are used. These values are combined with the Rainfall dataset.

C. Data pre-processing:-

For initial analysis purposes Standardization of feature values was performed using Standard Scaler and categorical features were label encoded using Label Encoder before using data for training the model.

IV. Methodology:-



Fig 2. Modular Diagram for the proposed system

A. Understanding correlation of climatic factors on the spices yield

A correlation matrix is a table showing correlation coefficients between variables. Each cell in the table shows the correlation between two variables. A correlation matrix is used to summarize data, as an input into a more advanced analysis, and as a diagnostic for advanced analyses. According to Fig 3, all the climatic factors such as Rainfall,Temperature,Air Humidity as well as the soil parameters such as Soil moisture and Soil pH level were taken into consideration.

Temperature has a good correlation with Crop Year and Area.Humidity has a good correlation with Area and Temperature.Soil moisture has a good correlation with Year,Area and Temperature.This detailed analysis helped us to predict the climatic factors for the forthcoming years.

The final correlation plot i.e. Fig 4 has the climatic factors which have a good correlation with the production. The factors are Temperature, Humidity and Soil moisture.



Fig 3. Correlation matrix for all climatic factors



 Fig 4. Correlation plot with selected climatic

 GSJ© 2021
 factors

B. Prediction Model for Yield

The dataset is split into 75% train data and 25% test data. The algorithms used were Linear Regressor, Random Forest Regressor, Stochastic Gradient Descent (SGD) Regressor and KNN. These algorithms were applied on scaled as well as non scaled data. On plotting box plots as in Fig 5. and Fig 6., the presence of outliers was noticed which reduced the accuracy of the above mentioned algorithms. Since the area provided for cultivation varied from State to State the production also varied drastically. Training the model on the entire dataset did not provide accurate results. Therefore



Fig 5. Production outliers

V. RESULT:-

This paper demonstrates the use of various regression algorithms Random Forest Regressor,Stochastic Gradient Descent(SGD), Linear Regression, Keras, KNN for predicting spice yield in a region.A comparative study was performed using five different regression algorithms to enhance accuracy.The accuracies of



Fig 7. Mean absolute error bar graph

we decided to train the model based on the user input. The user input includes the parameters:-

- State
- District
- Area
- Crop
- Year
- Season

The dataset would be segregated according to the above parameters provided by the user. The model would receive the training data as per the above parameters inputted by the user. The climatic factors for the desired year will be predicted and fed to the model.



Fig 6. Area outliers

different models were compared after building the models.Random Forest Classifier Algorithm tops the list of the lowest mean absolute error. Fig 7 shows the bar graph representation of mean absolute score of the different algorithms used to analyze the yield.

	Algorithm used	CropYield Prediction r2 score	Comments
	Random Forest Regressor	0.6	Outliers reducing accuracy
	Stochastic Gradient Descent Regressor	0.237	Overfitting
	KernelRidge Regressor	0.8	Handles the outliers and works well on medium sized datasets.
	KNN	0.238	Sensitive to scale of data and irrelevant features

Fig 8. R2_score comparison



Fig 11. Dry Chilly





Fig 13. Yield Prediction of Coriander GSJ© 2021 www.globalscientificjournal.com

- Fig 13 shows the prediction of Production of Coriander for the state Bihar, district wise using Kernel Ridge
- For the WPI section, the model predicts WPI values from June 2021 to May 2022. Fig 9 12, show the variation of predicted WPI values for the crops Black Pepper, Turmeric, Cumin and Chilly, The red constant line in the above plots indicate current prices of the specified spices.

VI. FUTURE SCOPE:-

- 1. Use of Weather APIs in order to get real-time climatic data for prediction purposes.
- 2. Creating a system wherein the farmer can Login as well as Register and also see frequently asked questions.
- 3. Embedding a chatbot for farmers in multiple languages.

VII. CONCLUSION:-

The proposed framework gives a brief idea about the spices which are most productive in a particular region based on the input parameters of the user. As the framework drills down every conceivable harvest, it helps the cultivator in the dynamic of which yield to develop at which season. Later IOT devices can also be used with this model in order to receive accurate data like soil pH levels, soil moisture, temperature and humidity. The various algorithms used give us an insight into the features which are most useful in predicting the yield. The future endeavour would be to include more attributes which would help in improving the yield results.

VIII. REFERENCES:-

[1] Rituraj Tiwari, ET Bureau, "India's farm exports rise 23.24% in March-June 2020 despite pandemic" ET Prime ,Aug 18, 2020. [Online], Available: https://economictimes.indiatimes.com

FE Bureau,"COVID-19 impact: Indian [2] spices rise in demand for immunity properties, exports up 34% in June", July 18, 2020, [Online], Available:https://www.businesstoday.in/

PB Jayakumar, "India's spices exports grow [3] 19% during the first half of FY21 ", December 23, 2020 .[Online], Available: https://www.financialexpress.com

B Sasikumar, "Cardamom - The third most [4] expensive spice in the world", FBNews, 17 August, 2015[Online], Available: http://www.fnbnews.com

Mordor Intelligence, "DRY CHILLIES [5] MARKET - GROWTH, TRENDS, COVID-19 IMPACT, AND FORECASTS (2021 - 2026)", 2021. [Online]. Available:

https://www.mordorintelligence.com/industry-reports /dry-chillies-market

[6] SML Venkata Narasimhamurthy, AVS random forest algorithm", International journal for

P.Priya, U.Muthaiah, M.Balamurugan, [7] "Predicting yield of the crop using machine learning Algorithm", IJESRT et al., 7(4): April-2018, pp 2277-2284

Filippi, P., Jones, E. J., Wimalathunge, N. [8] S., Somarathna, P. D., Pozza, L. E., Ugbaje, S. U., ... & Bishop, T. F. (2019). An approach to forecast grain crop yield using multi-layered, multifarm data sets and machine learning. Precision Agriculture, 1-15.

Tamsekar, P., Deshmukh, N., [9] Bhalchandra, P., Kulkarni, G., Hambarde, K., & Husen, S. (2019). Comparative Analysis of Supervised Machine Learning Algorithms for GIS-Based Crop Selection Prediction Model. In Computing and Network Sustainability (pp. 309-314)

[10] Saranya C P, Guru Murthy B, Karuppasamy M, Sunmathi M, Shree Sakthi Keerthna S (2020) "A SURVEY ON CROP YIELD PREDICTION USING MACHINE LEARNING ALGORITHMS".

pavan Kumar, "Rice crop yield forecasting using research in Applied science and Engineering technology, vol.5, Issue 10, october 2017, Pp:1220-1225