CROP YEILD AND PRICE PREDICTION USING MACHINE LEARNING

By

19BCE7567 – Yandra Hemanth
19BCE7513 – Akella Sivasai Atchyut

Guided By
Dr. D. Khasim Vali

PROBLEM STATEMENT

Based on prior crop yield, nutrients supplied, fertilizers used, climatic conditions to achieve high crop yield and predicting the best crops that yield the highest value for the upcoming cultivating season.

AGRICULTURE IN INDIA

- Agriculture is called the backbone of Indian economy. India’s agriculture is composed of many crops like cereals, pulses, fruits, vegetables, coffee, tea, oilseeds, spices etc., and also non-food crops like rubber, jute, cotton etc.
- Agriculture plays a very key role in the global economy of the country. Due to the increase in population, there is a constant need for the agricultural systems to improve the productivity of the crops and grow more crops.
- For many farmers, the biggest worry is the price fluctuation of the crop. Due to unstable prices, farmers are never able to plan a definite production pattern. This problem is highly prevalent in crops like tomatoes that have very limited shelf time.

PRODUCTIVITY CHALLENGES IN INDIA'S AGRICULTURAL SECTOR

- **Farm Holdings**
  As farms are divided among family members, average farm size today is half of what it was 40 years ago, as farms are divided among families.
- **Irrigation**
  Water use efficiency is very low. Unsustainable practices such as flooding or canals and pumping ground water resources are depleting country's aquifers.
- **Indian Monsoon**
  The country is faced with prospect of declining rainfall during the monsoon. India' prime growing season for rain fed crops
- **Subsidy**
  Govt grants subsidy to farmers for fertilizer, electricity and irrigation Areas receiving the highest subsidy regularly underperform those with lower subsidy
- **Cold Storage and Food**
  Harvest and post-harvest economic losses occur in some sectors depending on the demand for the crop

- **Crop Pattern, Fertilizer and Soil Nutrients**
  Nearly 60% of the land is at risk because of fertilizer misuse, poor cropping pattern and soil nutrient deficiencies.

AI IN AGRICULTURE

- Artificial Intelligence in agriculture has brought an agricultural revolution. This technology has protected the crop yield from various factors like climate changes, population growth, employment issues, and food security problems.
- The use of AI in Agriculture allowed us to various applications of n agriculture such as for irrigation, weeding, spraying with the help of sensors and other means embedded in robots and drones.
- Artificial intelligence makes it possible for farmers to assemble a large amount of data from the government as well as public websites, analyze all of it and provide farmers with solutions to many ambiguous issues as well as it provides us with a smarter way of irrigation which results in higher yield to the farmers.
- Artificial intelligence powered chatbots, along with machine learning techniques has enabled us to understand natural language and interact with users in away more personalized way. They are mainly equipped for retail, travel, media, and agriculture has used this facility by assisting the farmers to receive answers to their unanswered questions, for giving advice to them and providing various recommendations also.
- The use of AI in Agriculture will allow us to Images of the plant flow into an artificial intelligence algorithm that predicts precisely how long it will take for the blossom to become a ripe tomato ready for picking, packing
- From detecting pests to predicting what crops will deliver the best returns, artificial intelligence can help humanity confront one of its biggest challenges: feeding an additional 2 billion people by 2050, even as climate change disrupts growing seasons, turns arable land into deserts, and floods once-fertile deltas with seawater.

CROP YIELD PREDICTION

- The main aim of this project is to help farmers to decide on what to grow and when to grow.

- The Agricultural yield primarily depends on weather conditions (rain, temperature, etc), pesticides. Accurate information about the history of crop yield is important for making decisions related to agricultural risk management and future predictions.
- We eat a lot of corn, wheat, rice, and other simple crops. In this project, machine learning methods are applied to predict various crops.

HOW THE PROJECT WORKS?

First, we take required inputs(attributes) from the farmers and calculate the yield for different crops, and then we can choose the best yield-giving crop.

LITERATURE SURVEY

1. Machine learning approach for forecasting crop yield based on parameters of climate. The paper provided in International Conference on Computer Communication and Informatics (ICCCI). The papers is based on parameters like rainfall, maximum and minimum temperature, potential evapotranspiration, cloud cover, wet day frequency. C4.5 algorithm is used to produce the most influencing climatic parameter on the crop yields of selected crops in selected districts of Madhya Pradesh. The paper is implemented using Decision Tree.
2. Crop production Ensemble Machine Learning model for prediction. International Journal of Computer Science and Software Engineering (IJCSSE). The parameters considered in this paper are average temperature, cloud cover, vapor pressure, wet day frequency, evapotranspiration. In this paper, AdaNaive and AdaSVM are the proposed ensemble model used to project the crop production over a time period. Implementation done using AdaSVM and AdaNaive. AdaBoost increases efficiency of SVM and Naive Bayes algorithm.
3. Random Forests for Global and Regional Crop Yield Predictions, Institute on the Environment, University of Minnesota, United States of America. The paper considered parameters like climate, soil, photoperiods, water supply and fertilization data. The generated outputs show that Random Forests is an effective and different machine-learning method for crop yield predictions at regional and global scales for its high accuracy. The Paper is Implemented using k-nearest neighbors, Support Vector Regression (SVR), Random Forests.

4. Applications of Machine Learning Techniques in Agricultural Crop Production. Indian Journal of Science and Technology, Vol 9(38). From GPS based color images, model is provided as an intensified indistinct cluster analysis for classifying plants, soil and residue regions of interest. The paper includes various parameters which can help the crop yield for better enhancement and ratio of the yield can be increased during cultivation. The paper is implemented using Artificial Neural Networks, Time series analysis and Decision Trees.
5. Predicting yield of the crop using machine learning algorithm. International Journal of Engineering Science Research Technology. The paper focuses on predicting the yield of the crop based on the existing data. Real data of Tamil Nadu were used for building the models and the models were tested with samples. The factors considered are rainfall, Kharif and Rabi seasons, maximum temperatures, precipitation etc.
6. Crop Prediction System using Machine Learning Algorithms, International Research Journal of Engineering and Technology, Vol 7(2). The paper focuses on making use of classification algorithms to improvise the crop yields. The paper is implemented using Decision Tree Classifier, Naïve Bayes Classifier and KNN Classifier.

TENTATIVE METHODOLOGY

- Problem Identification
- Data Collection
- Dataset Processing
- Data Visualization to gain insights
- Getting the data ready for training
- Training the machine learning models
- Finalizing the model that gives best crop yield
- Relating the crop yield with current and probable future market prices
- Creating a farmer friendly UI for the project

DATASET

The dataset is collected from official and trusted resources like FAO (Food and Agricultural Organization of the United Nations) and Indian Government's data resources like ODG (Open Government Data).

Multiple datasets are collected on the following categories:

- Crop yield
- Temperature data
- Fertilizers by nutrients containing Nitrogen, Potash, Phosphate
- Pesticides
- Quantity Produced
- Area Irrigated
- Area Harvested
- Land Used
- Synthetic Fertilizers

PROCESSING DATASETS

- All the collected datasets have the data related to 80 different crops that were grown in India spanning over 30 years (from 1990 to 2019).
- The columns in each dataset are checked thoroughly and unwanted features are removed.
- Redundant features are removed by keeping the important one out of them
- All the datasets are finally merged into a bigger dataset based on the crop items and the year of production.
- This new dataset is used for the further process.

SELECTING AND TRAINING THE MACHINE LEARNING MODEL

- From the problem and the dataset, we can infer that this is a regression problem.
- Various regression algorithms are selected and trained to find the best model out of them.
- The models trained in this project are
  1. Linear Regressor

2. Decision Tree Regressor
3. Random Forest Regressor
4. Gradient Boosting Regressor
5. Linear Support Vector Regressor (SVR)

TRAINING THROUGH CROSS VALIDATION

- Scikit-Learn's K-fold cross-validation feature.
- It randomly splits the training set into k distinct subsets called folds, then it trains and evaluates the model k times, picking a different fold for evaluation every time and training on the other (k-1) folds.
- In the current project we considered k = 10

EVALUATION METRICS

- When it comes to regression algorithms, accuracy of the model doesn't work.
- It is because we cannot compare between real values which extend infinitely between the limits.
- Hence for the regression algorithms, we try to find the error that a model makes.
- Root Mean Squared Error is one such evaluation metric, which calculates the error made by the model on all instances.

ROOT MEAN SQUARED ERROR

$$\mathrm{RMSE}(\mathbf{X}, h) = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(h\left(\mathbf{x}^{(i)}\right) - y^{(i)}\right)^2}$$

- R2 Score helps in determining how better the model can perform by calculating the proportion of variance in the dependent variable that is predictable from the independent variables.

R2 SCORE

$$R^2 = 1 - \frac{RSS}{TSS}$$

FUTURE POSSIBILITIES

- This project depends on the input given by the farmer based on the current price of the crop. This can be improved by collecting data on crop prices across the country.
- This project can be specialized regionally (for example, statewide crop yield prediction) depending on the factors prevailing in that region.
- More climatic factors like rainfall, precipitation, wind speed, humidity, etc can also be taken into consideration in order to improve the model's efficiency in terms of climatic conditions.
- Data related to the soil present in the crop can also be considered for more insights by examining the soil through other artificial intelligence techniques.
- Considering the effects of the natural disasters might help the farmers in exceptional environmental conditions.
- More crops can be added to the list of the crops in order to increase the reach of the project to more farmers.
- The application's UI can be made more farmer-friendly by supporting regional languages.

REFERENCES AND BIBLIOGRAPHY

- S. Veenadhari, B. Misra and C. Singh, "Machine learning approach for forecasting crop yield based on climatic parameters," 2014 International Conference on Computer Communication and Informatics, Coimbatore, 2014, pp. 1-5, doi: 10.1109/ICCCI.2014.6921718.
- Balakrishnan, N., Muthukumarasamy, G.: Crop production-ensemble machine learning model for prediction. Int. J. Comput. Sci. Softw. Eng. **5**(7), 148–153 (2016)
- Jeong JH, Resop JP, Mueller ND, Fleisher DH, Yun K, Butler EE, Timlin DJ, Shim KM, Gerber JS, Reddy VR, Kim SH. Random Forests for Global and Regional Crop Yield Predictions. PLoS One. 2016 Jun 3;11(6):e0156571. doi: 10.1371/journal.pone.0156571. PMID: 27257967; PMCID: PMC4892571.
- Mishra, Subhadra & Mishra, Debahuti & Santra, Gour. (2016). Applications of Machine Learning Techniques in Agricultural Crop Production: A Review Paper. Indian Journal of Science and Technology. 9. 10.17485/ijst/2016/v9i38/95032.

- Priya, P., Muthaiah, U., & Balamurugan, M. (2018). Predicting yield of the crop using machine learning algorithm. *International Journal of Engineering Sciences & Research Technology*, *7*(1), 1-7.
- Pavan P., Virendra P., Shrikhant K., *Crop Prediction System using Machine Learning Algorithms,* International Research Journal of Engineering and Technology (IRJET), 7(2), 2020 Feb.
- Aurelien Geron, *Hands-on Machine Learning with Scikit-Learn, Keras and TensorFlow,* 2nd Edition, O'Reilly.
- David Freedman, Robert Pisani, Roger Purves, *Statistics,* 4th Edition, Viva Books Pvt ltd.
- Wes McKinney, *Python for Data Analysis*, 2nd Edition, O'Reilly.
- Jake VanderPlas, *Python Data Science Handbook*, O'Reilly.
- Peter Harrington, *Machine Learning in action*, Manning Publications
- Samprit Chatterjee, Ali S. Hadi, *Regression Analysis by Example*, 5th Edition, Wiley.