Crop recommendation using Weather Data

Rudra Shah,

Department of CSE,
PDEU, Gandhinagar, Gujarat

Vinit Prajapati,

Department of CSE,
PDEU, Gandhinagar, Gujarat

Hrishikesh Kalola,

Department of CSE,
PDEU, Gandhinagar, Gujarat

Dr. Archana Nigam,
Department of CSE,
PDEU, Gandhinagar,
Gujarat.

Abstract— This project harnesses big data analytics to predict crop yield based on key environmental factors such as temperature, humidity, and rainfall. Leveraging a neural network architecture, the system accurately forecasts the most suitable crops for a given set of environmental conditions. Furthermore, the project extends its capabilities by integrating nutrient recommendation functionalities and providing tailored suggestions for nitrogen, phosphorus, and potassium (NPK) values based on the predicted crop types. The methodology involves data preprocessing, feature scaling, and the development of an artificial neural network (ANN) model for prediction. The model is trained on a dataset comprising historical environmental data and corresponding crop yields. Upon validation, the system demonstrates high accuracy in predicting optimal crop choices. Additionally, the integration of nutrient recommendations enhances its utility, enabling farmers to make informed decisions to maximize crop productivity and sustainability. This project represents a significant advancement in agricultural technology, empowering farmers with data-driven insights to optimize yield and resource utilization.

INTRODUCTION

Agriculture forms the backbone of global food production, making it imperative to optimize crop yield while ensuring sustainable resource management. Traditional farming practices often rely on intuition and historical knowledge, lacking precision and efficiency. With the advent of big data technologies, there is an opportunity to revolutionize agriculture by leveraging vast datasets to make informed decisions.

This project addresses the need for a data-driven approach to crop prediction and nutrient management. By analyzing historical environmental data including temperature, humidity, and rainfall, we aim to predict the most suitable crops for specific conditions. Additionally, we introduce a novel feature of nutrient recommendation, tailoring nitrogen, phosphorus, and potassium (NPK) values based on predicted crop types.

The methodology involves several key steps. First, we preprocess the dataset, ensuring data quality and relevance. Next, we employ feature scaling techniques to normalize the data and prepare it for modeling. Subsequently, we develop an artificial neural network (ANN) model capable of predicting crop types based on environmental parameters. The model is trained on a diverse dataset encompassing various climatic conditions and crop varieties.

Upon model validation, we evaluate its performance in accurately predicting crop choices. Furthermore, we demonstrate the effectiveness of the nutrient recommendation system, providing personalized suggestions to enhance crop yield and quality. The integration of these functionalities represents a significant advancement in agricultural technology, empowering farmers with actionable insights to optimize their farming practices.

DATASET DESCRIPTION

Crop Dataset Description:

The dataset is compiled by augmenting existing datasets on rainfall, climate, and fertilizer data in India. It provides essential parameters influencing crop growth and productivity, facilitating the development of tailored predictive models for crop recommendations.

Data Fields:

N (Nitrogen): *Ratio of Nitrogen content in soil, vital for plant growth.*

P (Phosphorous): *Ratio of Phosphorous content in soil, crucial for root development.*

K (Potassium): *Ratio of Potassium content in soil, essential for plant resilience.*

Temperature: *Ambient temperature in degrees Celsius, influencing crop growth rates.*

Humidity: *Relative humidity levels (%), impacting plant moisture management.*

pH: *Soil pH value, affecting nutrient availability.*

Rainfall: *Amount of rainfall in mm, crucial for soil moisture and irrigation.*

City Weather Dataset Description:

***Context**:*

This dataset provides historical weather data for the top 8 Indian cities based on population. It was created to address the need for comprehensive weather data within the community, spanning over a decade from 2009 to 2020.

Content:

The dataset was obtained using the worldweatheronline.com API and the wwo_hist package, offering hourly weather data for each city. With over 10 years of data per city, it enables analysis of weather patterns and trends, including the effects of global warming. However, it's essential to note that the data's accuracy cannot be guaranteed.

Acknowledgements:

The weather data is sourced from worldweatheronline.com and extracted using their API, acknowledging their contribution to making this dataset available.

Inspiration:

This dataset is valuable for predicting weather conditions for upcoming days, weeks, or months due to its extensive historical coverage. Additionally, it serves as a foundation for visualizations, aiding in understanding the impact of global warming on various weather parameters such as precipitation, humidity, and temperature.

Columns:

1. *date_time*
2. *maxtempC*
3. *mintempC*
4. *totalSnow_cm*
5. *sunHour*
6. *uvIndex*
7. *uvIndex.1*
8. *moon_illumination*
9. *moonrise*
10. *moonset*
11. *sunrise*
12. *sunset*
13. *DewPointC*
14. *FeelsLikeC*
15. *HeatIndexC*
16. *WindChillC*
17. *WindGustKmph*
18. *cloudcover*
19. *humidity*
20. *precipMM*
21. *pressure*
22. *tempC*
23. *visibility*
24. *winddirDegree*
25. *windspeedKmph*

METHODOLOGIES

1. Data Acquisition and Preprocessing:

Data Collection: The dataset containing information on crops and environmental factors such as temperature and humidity is imported from 'Crop_recommendation.csv'.
Feature Selection: The features relevant to the prediction task, namely 'temperature' and 'humidity', are extracted from the dataset.
Target Encoding: The target variable 'label' is one-hot encoded using sklearn's OneHotEncoder to prepare it for multi-class classification.
Data Splitting: The dataset is divided into training and testing sets using a 80-20 split ratio. This ensures that the model's performance can be evaluated on unseen data.
Feature Scaling: The features in the dataset are scaled using the StandardScaler from sklearn to normalize them, enhancing the model's convergence during training.

2. Model Development:

Artificial Neural Network (ANN) Architecture: An ANN model is constructed using TensorFlow's Keras API. The model architecture consists of an input layer, three hidden layers, and an output layer. Each hidden layer is comprised of 22 neurons with ReLU activation functions.
Model Compilation: The ANN model is compiled using the 'Adam' optimizer and 'categorical_crossentropy' loss function, suitable for multi-class classification tasks. Additionally, the model's accuracy metric is tracked during training.

3. Model Training:

The ANN model is trained on the training dataset using a batch size of 32 and 250 epochs. During training, the model learns to minimize the loss function and improve its predictive performance.

4. Model Evaluation:

Prediction and Evaluation: The trained model is used to predict the crop labels for the test dataset. The predictions are then evaluated using metrics such as accuracy, precision, recall, and F1 score. These metrics provide insights into the model's performance and its ability to correctly classify crops based on environmental factors.

5. Prediction for Different Locations:

The trained model is applied to predict the suitable crops for various locations including Delhi, Bengaluru, Bombay, Pune, Hyderabad, Jaipur, Kanpur, and Nagpur. The environmental data for each location is loaded from separate CSV files and fed into the model to generate crop predictions.

6. Dataset Export:

The predicted crop labels for each location are appended to their respective datasets. These modified datasets are then saved as CSV files for further analysis or application.

7. Data Packaging:

The generated CSV files are compiled into a single zip file named 'datasets.zip' for easy sharing and distribution.

8. Download and Conclusion:

The zip file containing the datasets is made available for download using the Google Colab environment. This concludes the methodology section, summarizing the steps involved in data preprocessing, model development, training, evaluation, and prediction for crop recommendation based on environmental factors.

Conclusion:

In this project, we developed an Artificial Neural Network (ANN) based system for crop recommendation utilizing environmental factors such as temperature and humidity. The methodology involved several key steps including data preprocessing, model development, training, evaluation, and prediction for multiple locations.

Through rigorous data preprocessing, we ensured the dataset's compatibility with the model training pipeline by encoding categorical variables and standardizing numerical features. The ANN model architecture was designed with three hidden layers, each consisting of 22 neurons with rectified linear unit (ReLU) activation functions. The model was trained using the Adam optimizer and categorical crossentropy loss function, achieving satisfactory performance metrics on the test dataset, including accuracy, precision, recall, and F1 score.

Furthermore, the trained model was applied to predict suitable crop recommendations for various locations including Delhi, Bengaluru, Bombay, Pune, Hyderabad, Jaipur, Kanpur, and Nagpur. The environmental data for each location was fed into the model, generating crop predictions that can assist farmers and agricultural stakeholders in making informed decisions regarding crop selection based on prevailing environmental *conditions.*

The predicted crop labels were appended to their respective datasets and saved as CSV files for further analysis or application. Additionally, the datasets were packaged into a single zip file for easy sharing and distribution.

In conclusion, the developed ANN-based crop recommendation system demonstrates promising potential for enhancing agricultural practices by leveraging machine learning techniques to optimize crop selection and improve yield outcomes in diverse environmental conditions. As future work, further refinement and validation of the model could be pursued, along with the integration of additional environmental variables and data sources to enhance prediction accuracy and robustness.

References:

1. ***Weather Dataset -*** *https://data.opencity.in/dataset/daily-temperature-70-years-data-for-major-indian-cities* *https://www.kaggle.com/datasets/hiteshsoneji/historical-weather-data-for-indian-cities*
2. ***Crop dataset -*** *https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset*
3. ***Comparative study -*** *https://iopscience.iop.org/article/10.1088/1742-6596/2161/1/012033/pdf*
4. ***Crop-Weather relation -*** *https://krishi.icar.gov.in/jspui/bitstream/123456789/32973/1/Crop-weather%20relationships.%20Cropping%20system%20management.pdf*
5. ***Recommendations using GNB (Gaussian Naïve Bayes) -*** *https://www.researchgate.net/publication/370056714_Crop_Recommendation_System*
6. ***Recommendation using GCN (graph convolution network) -*** *https://www.mdpi.com/2673-4591/58/1/97*
7. ***Crop prediction using different features -*** *https://www.researchgate.net/publication/346627389_Crop_Recommendation_System*