DATA-DRIVEN INSIGHTS FOR AGRICULTURE

Intellihack_CyperZ

Task 01

Dhanushi Dewmindi
Amanda Hansamali

WEATHER PREDICTION MODEL REPORT

Contents

1. Introduction

Accurate weather predictions are crucial for farmers to effectively plan irrigation, planting, and harvesting. Traditional weather forecasts often lack the granularity needed for hyper-local predictions. This report details the development of a machine learning model that predicts the probability of rain based on historical weather data.

2. Dataset Overview

The dataset consists of 300 daily weather observations with the following features:

- **avg_temperature**: Average temperature in °C
- **humidity**: Humidity in percentage
- **avg_wind_speed**: Average wind speed in km/h
- **rain_or_not**: Binary target variable (1 = rain, 0 = no rain)
- **date**: Date of observation

3. Data Preprocessing

3.1 Handling Missing Values

- The dataset was inspected for missing values and cleaned using SimpleImputer with the mean strategy.
- Non-numeric or improperly formatted values were converted to numeric using pd.to_numeric with error coercion.
- Rows with invalid or missing essential data were dropped.
- After cleaning, the dataset was verified to ensure it was not empty and contained sufficient samples for training.

3.2 Data Formatting

- The date column was converted to a proper datetime format.
- Non-numeric columns were excluded from the correlation matrix to avoid processing errors.

4. Exploratory Data Analysis (EDA)

4.1 Correlation Matrix

- A heatmap was generated to visualize correlations between numerical features.
- No strong linear correlations were observed, indicating the need for a robust model to capture non-linear patterns.

4.2 Target Variable Distribution

- The distribution of rain vs. no rain was visualized using a count plot, showing a balanced dataset which is ideal for model training.

5. Model Development

5.1 Model Selection

- A RandomForestClassifier was chosen for its robustness and ability to handle both linear and non-linear relationships.

5.2 Model Training

- The dataset was split into training (80%) and testing (20%) sets using train_test_split.
- Feature scaling was applied using StandardScaler to normalize input features.

5.3 Model Evaluation

- The initial model achieved an accuracy score of **X%**.
- The confusion matrix and classification report showed balanced precision and recall for both classes.

5.4 Hyperparameter Tuning

- Grid search (GridSearchCV) was performed to optimize model parameters including:
    - n_estimators: Number of trees in the forest
    - max_depth: Maximum depth of the trees
    - min_samples_split: Minimum samples required to split a node
- The best parameters found were:

{'n_estimators': 100, 'max_depth': None, 'min_samples_split': 2}

- The optimized model achieved an accuracy score of **Y%** on the test set.

6. Prediction Results

6.1 Future 21-Day Rain Probability

- The model generated probability predictions for rain over the next 21 days.
- These predictions offer actionable insights for farmers to make informed decisions.

7. Conclusions

- The machine learning model provides reliable rain probability forecasts, with an accuracy improvement of **Z%** after hyperparameter tuning.
- Future enhancements could include incorporating additional features such as cloud cover and pressure data, if available.

8. Recommendations

- Integrate this model with an API for real-time data updates.
- Deploy the model within a smart agriculture platform to provide automated alerts to farmers.

9. Appendix

- Full code and detailed analysis are included in the accompanying Jupyter Notebook.