Predicting the economic impact of climate and agricultural factors using Machine Learning

Introduction:

The agricultural sector is highly sensitive to climatic variations and soil conditions. Understanding how these factors affect the economic outcomes of farming activities is critical for planning, risk management, and policy decisions. This project aimed to predict the economic impact (in million USD) of climate, soil, and agricultural practices using machine learning models.

Data Description:

The dataset included 10,000 observations across multiple countries and regions, with features such as:

- Climate variables: Average temperature (°C), total precipitation (mm), CO_2 emissions (MT), extreme weather events.
- Soil and agricultural inputs: Soil health index, irrigation access (%), fertilizer and pesticide usage (kg/ha), crop type, crop yield (MT/ha).
- Adaptation strategies adopted by farmers (e.g., drought-resistant crops, water management).
- Target variable: Economic impact (million USD).

Methodology:

Exploratory Data Analysis (EDA):

- Examined distributions of climate, soil, and agricultural variables using histograms and facet grids.
- Investigated economic impact across countries, regions, and adaptation strategies using box plots and bar charts.
- Explored trends of crop type and crop yields by country over time.

Data Preprocessing:

- Selected the most important features based on correlation analysis and feature importance from Random Forest model.
- Handled categorical variables and ensured consistency in data types for modeling.

Model Training and Validation:

- Implemented three model Decision Tree, Random Forest, and XGBoost regressors.
- To ensure robustness conducted hyperparameter tuning using RandomizedSearchCV with 5-fold cross-validation.

- Evaluated models based on R^2, RMSE, and MAE on both training and test sets.

Model Evaluation and hyperparameter tuning:

Among the models tested, the Random Forest regressor consistently outperformed the Decision Tree and XGBoost models. After tuning, it achieved:

- Cross-validated R^2**:** 0.49 (average across folds)
- Test set R^2: 0.58(after tuning)

Feature importance analysis highlighted crop yield, average temperature, irrigation access, and soil health as the strongest predictors of economic impact. Residual diagnostics showed no systematic bias, indicating a reliable fit.

Results and Insights:

The findings provide strong evidence that climate, soil, and farming practices significantly affect agricultural economic outcomes. Specifically:

- Adaptation strategies such as drought-resistant crops and improved water management were associated with lower economic losses.
- The predictive model provides actionable insights for policymakers, agricultural companies, and farmers, helping them allocate resources efficiently and mitigate risks.

Model Performance and Limitations:

The Random Forest model achieved an R^2 of 0.58 on the test set, indicating that it explains about 58% of the variation in economic impact. While this is a moderate performance, several factors may limit predictive accuracy:

- High variability in economic impact**:** Economic outcomes are influenced not only by climate and soil factors but also by external elements such as market fluctuations, policy changes, and global trade, which are not included in the dataset.
- Complex interactions**:** Nonlinear interactions between climate, soil, and management practice may not be fully captured by the current feature set.

Conclusion:

This project illustrates how machine learning can quantify the economic impact of climate and agricultural factors. The Random Forest model, combined with careful feature selection and tuning, provides a robust tool for predicting economic outcomes and understanding the drivers of financial risk in agriculture.

Appendix: The complete code available at: https://github.com/ismat210/economic-impact-of-climate-and-agricultural-factors-