Crop Yeild Prediction

Abstract

Crop Yield Prediction Project aims to forecast crop yields using machine learning models. The project utilizes data obtained from Kaggle, specifically focusing on the usage of pesticides and its impact on crop yields. Two machine learning models, Linear Regression and Support Vector Regression (SVR), are implemented and evaluated for prediction accuracy. Additionally, a Random Forest Regressor is employed for comparison.

Introduction

The project addresses the crucial task of predicting crop yields, a key factor in agricultural planning. Understanding the relationship between pesticide use and crop yields is vital for optimizing agricultural practices. The hypotheses involve exploring whether machine learning models can effectively predict crop yields based on historical data.

Data

The dataset, sourced from Kaggle, comprises information on pesticide use in various regions, including details such as Domain, Area, Element, Item, Year, Unit, and Value. The data undergoes preprocessing, including cleaning and transformation, to prepare it for model training.

Methods

Linear Regression

A Linear Regression model is built and trained using the dataset. The model's performance is assessed using the R^2 score, indicating its ability to explain the variance in the target variable.

Support Vector Regression (SVR)

SVR, a type of Support Vector Machine (SVM), is employed for regression tasks. The model is trained on scaled data, and predictions are made on the test set. Evaluation metrics include Mean Squared Error (MSE) and R-squared (R^2).

Random Forest Regressor

The Random Forest Regressor was initialized, trained on the dataset, and predictions were made on the test set. The model's performance was evaluated using the Mean Squared Error (MSE) and R-squared (R^2) metrics.

The Random Forest Regressor achieved impressive results:

Random Forest Mean Squared Error: 0.00047
Random Forest R-squared: 0.968
These metrics indicate a high level of accuracy in predicting crop yields based on pesticide use. The low MSE suggests that the model's predictions are very close to the actual values. The R^2 score of approximately 0.968 further emphasizes the model's ability to explain about 96.8% of the variance in crop yields.

Results

Linear Regression

The Linear Regression model achieves an R^2 score of approximately 0.72 on the test set, indicating that 72% of the variance in crop yields can be explained by the features.

Support Vector Regression (SVR)

The SVR model demonstrates a Mean Squared Error of approximately 0.0057 and an R^2 score of around 0.617, suggesting a moderate level of predictive accuracy.

Random Forest Regressor

The performance of the Random Forest Regressor is yet to be presented.

Discussion

The three distinct machine learning models—Linear Regression, Support Vector Regression (SVR), and Random Forest Regressor—to predict crop yields based on pesticide use. Each model exhibited unique characteristics, and their performance was evaluated using key metrics.

Linear Regression

The Linear Regression model achieved an R^2 score of approximately 0.72 on the test set, indicating that 72% of the variance in crop yields could be explained by the features. While this is a respectable performance, it is essential to note that Linear Regression assumes a linear relationship between features and the target variable.

Support Vector Regression (SVR)

SVR, a more complex model, demonstrated a Mean Squared Error (MSE) of approximately 0.0057 and an R^2 score of around 0.617. This suggests a moderate level of predictive accuracy, and the model accounts for approximately 61.7% of the variance in crop yields. SVR is particularly useful when dealing with non-linear relationships.

Random Forest Regressor

The Random Forest Regressor showcased remarkable predictive accuracy with a very low Mean Squared Error (MSE) of 0.00047 and an impressive R^2 score of 0.968. This indicates that the Random Forest model explains about 96.8% of the variance in crop yields. The ensemble nature of Random Forest, combining multiple decision trees, likely contributed to its

superior performance.

Comparative Analysis

Comparing the models, Random Forest Regressor outperformed both Linear Regression and SVR in terms of accuracy. Its ability to capture complex relationships and handle non-linearity in the data makes it well-suited for this prediction task. Linear Regression and SVR, while still providing valuable insights, may have limitations in capturing intricate patterns present in the dataset.

Conclusion

In conclusion, the Random Forest Regressor, along with Linear Regression and Support Vector Regression (SVR), demonstrates its efficacy in predicting crop yields based on pesticide use. The comprehensive evaluation of these models suggests their potential for practical applications in agriculture. Future studies could delve deeper into feature importance and consider ensemble methods for even more robust predictions.