Multi trait Prediction System

In Wheat plant

Objectives:

Explore Relationships:

- Explore relationships among wheat plant traits

Phenotypic Prediction:

- Develop a model predicting observable traits from genetic data.

Feature Reduction:

- Simplify the dataset, reducing features for enhanced model efficiency.

Apply Random Forest for Predictions:

- Apply Random Forest to predict traits with the simplified dataset.

Model Optimization:

- Fine-tune the model for improved accuracy and generalization.

Project Deployment Readiness:

- Prepare the project for evaluation and potential deployment.

Steps:

1.Collected two datasets

Genotype data:

Genotype *is an organism's genetic instructions for traits.* ***Genotype data*** *comes from studying these instructions, using methods like reading DNA sequences*

Phenotype data:

Phenotype is the observable characteristics or traits of an organism, influenced by both genetic and environmental factors. Phenotype data includes information about the physical, biochemical, or behavioral traits of individuals.

2.Preprocessing:

Preprocess the data for machine learning, which may involve feature scaling, encoding categorical variables, or handling imbalanced classes.

3.Correlation Analysis:

Applying correlation to quantify the relationships between each phenotypic trait and single nucleotide polymorphism (SNP) values.

4.Data Merging:

Combined genotype and phenotype data, creating a unified dataset for analysis.(common values)

5.Backtracking

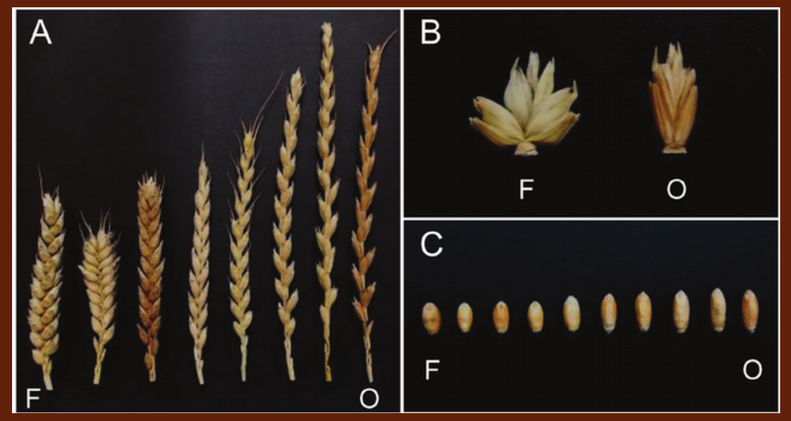
Applied backtracking techniques to reduce the dataset from thousands of columns to a more manageable 10-20 columns.

6.Model Selection:

Selected Random Forest for predicting traits because it combines the insights of multiple decision trees, making it effective in understanding complex patterns

7.Model Training:

Trained the Random Forest model using the prepared data to learn patterns and relationships.



8.Optimization:

Fine-tuned the model parameters for better accuracy and generalization.

9.Testing Data:

Assess the model's performance on a separate testing dataset no used during training. Evaluate metrics such as mean Squared error to gauge how well the model generalizes to new data

10. Flask Application for predictions

Create a Flask web application to serve predictions by using VSCode.

Built a Flask web app that links the trained model to the front end. Integrated the model into the Flask app to generate predictions using input data.

11.Frontend Development:

Designed and implemented the front end using HTML, CSS, and JavaScript. Created a user interface where users can input allele sequences.

Thank you!!!