**Regression Analysis on Iris\_Modified Dataset with Feature Engineering and Model Tuning**

This project performs regression analysis on the Iris\_Modified dataset (iris.csv) using various machine learning models. The objective is to predict the petal\_width based on other features after applying feature engineeriang techniques and optimizing model hyperparameters.

**Table of Contents**

**Project Overview** : This project aims to build and evaluate regression models to predict petal\_width using the Iris dataset. It includes preprocessing steps, feature engineering techniques, model selection, hyperparameter tuning, and evaluation based on regression-specific metrics.

**Dataset**: The dataset used in this project is iris.csv, which is a modified version of the classic Iris dataset. It contains the following features:

*sepal\_length*: Length of the sepal (numerical)

*sepal\_width*: Width of the sepal (numerical)

*petal\_length:* Length of the petal (numerical)

*petal\_width:* Width of the petal (target variable, numerical)

*species:* Species of the Iris flower (categorical)

**Feature Engineering**

**Missing Values Handling**

Imputation: Missing values in sepal\_width and petal\_width were imputed using custom methods.

**Feature Interaction**

Linear Interactions: Created interactions between petal\_length and sepal\_width.

Polynomial Interactions: Generated polynomial features for petal\_length / sepal\_width and petal\_width / species.

**Feature Selection**

Selected features: sepal\_length, sepal\_width, petal\_length, petal\_width, species.

**Modeling** **Approach**

**Regression Models**

Random Forest Regressor: Implemented with hyperparameter tuning.

Gradient Boosted Trees: Considered for potential implementation.

Decision Tree Regressor: Evaluated as a baseline model.

**Hyperparameter Optimization**

Grid Search: Used to optimize parameters such as max\_depth, min\_samples\_leaf, and n\_estimators for each model.

**Evaluation Metrics**

Models are evaluated based on:

R-squared (R^2): Measure of how well the model captures the variance in the target variable.

Mean Squared Error (MSE): Indicates the average squared difference between actual and predicted values.

**Output**

**Best Parameters for RandomForestRegressor**

max\_depth: 20

min\_samples\_leaf: 10

n\_estimators: 10

**Best Score for RandomForestRegressor**

R-squared (R^2): 0.9217