

## Assignment: Algerian Forest Fire Data Analysis using Multiple Linear Regression

**Objective:** The objective of this assignment is to analyze the FWI in the Algerian forest fire dataset using multiple linear regression, evaluate the model's performance, create pickle files, and test the model on unseen data.

### Instructions:

1. **Data Preparation:** a. Obtain the Algerian forest fire dataset, containing variables such as temperature, humidity, wind speed, location, and fire intensity. b. Clean the dataset, handling missing values and outliers if necessary. c. Split the dataset into a training set (80% of the data) and a test (unseen) set (20% of the data).
2. **Model Training:** a. Use multiple linear regression to model the relationship between the independent variables (temperature, humidity, wind speed) and the dependent variable (fire intensity). b. Fit the multiple linear regression model to the training set. c. Tune hyperparameters, if necessary, to improve the model's performance.
3. **Model Evaluation:** a. Evaluate the trained multiple linear regression model's performance on the test dataset. b. Calculate appropriate evaluation metrics such as mean squared error (MSE), mean absolute error (MAE), or coefficient of determination (R-squared). c. Interpret the results and assess how well the model predicts fire intensity.
4. **Pickling the Model:** a. Once you have a satisfactory multiple linear regression model, save it as a pickle file. b. Explain the purpose and benefits of pickling a model.
5. **Prediction on Unseen Data:** a. Load the pickled multiple linear regression model. b. Preprocess the unseen data using the same transformations applied to the training data. c. Use the loaded model to predict the fire intensity on the unseen data. d. Evaluate the model's performance on the unseen data using the same evaluation metrics as in Step 3.
6. **Conclusion:** Summarize the findings of your analysis on the Algerian forest fire data using multiple linear regression. Highlight the strengths and limitations of the model and provide recommendations for further improvements.

### Deliverables:

1. Jupyter Notebook or Python script containing the code for data preparation, model training, model evaluation, pickling the model, and prediction on unseen data.
2. Pickle file(s) containing the trained multiple linear regression model(s).
3. A conclusion section summarizing the analysis and providing recommendations.

**Note:** Provide clear and detailed explanations in your code and include appropriate comments to enhance understanding.

Create your assignment in any Python IDE notebook and upload it to GitHub & share that GitHub repository link through your dashboard. Make sure the repository is public.