Lab Task : 07

**Regression** models are essential tools in statistics and machine learning for understanding relationships between variables.

The dataset we use in the lab task is '*Alegerian Forest Fires Cleaned Dataset*'. To analysis the dataset we used different types of Regression model.

1) Linear Regression
2) Lasso Regression
3) Ridge Regression
4) Elasticnet Regression
5) Cross-Validation Lasso

**Linear Regression** : Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. It predicts outcomes by fitting a straight line through the data points, allowing for the estimation of how changes in the independent variable(s) affect the dependent variable.

**Lasso Regression :** Lasso (Least Absolute Shrinkage and Selection Operator) regression is a type of linear regression that includes a regularization term to penalize the absolute size of the coefficients. This method helps in reducing overfitting by shrinking some coefficients to zero, effectively performing variable selection.

**Cross-Validation Lasso:** Lasso regression combined with a cross-validation procedure to select the best value of λ.

**Ridge Regression** : Ridge regression also incorporates a regularization term, but it penalizes the sum of the squares of the coefficients instead of their absolute values. This approach helps to address multicollinearity among predictors by reducing their variance without eliminating any variables entirely.

**Elastic Net regression** : Elastic Net regression combines both Lasso and Ridge techniques by applying penalties based on both L1 (Lasso) and L2 (Ridge) norms. This method is particularly effective in situations where there are multiple correlated features, as it can select groups of variables while maintaining regularization.

**Why we used this model** ?

From this five regression model for '*Algerian Forest Fires Cleaned Dataset*'

We use it for 1) Linear Regression :Simplicity and We can inter-predict the data.

2) Lasso Regression : We can apply regularization and it prevents overfitting.

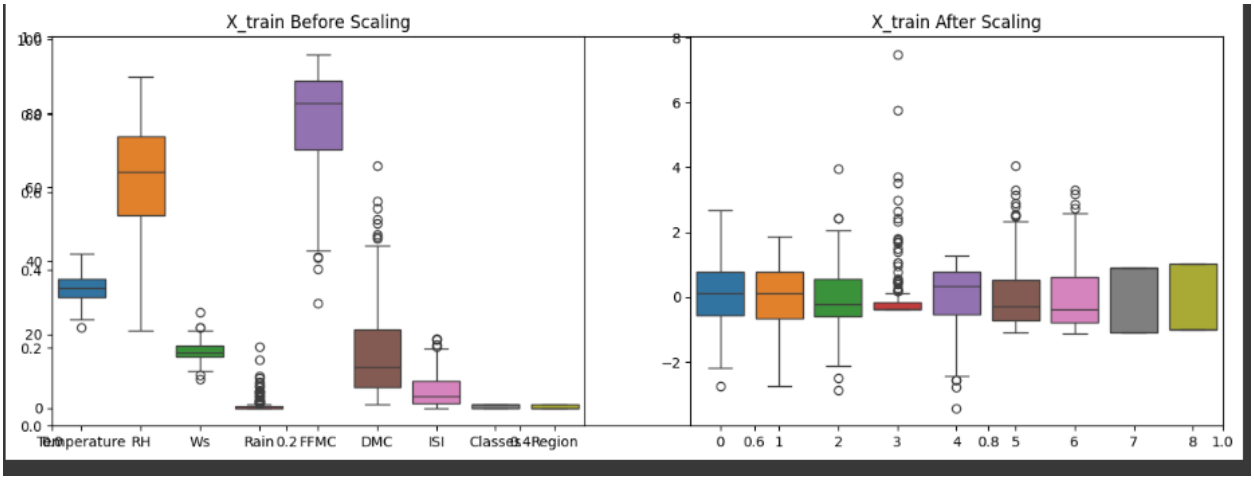
3)Ridge Regression: Handle collision and suitable for all variables

4)Cross Validation Lasso : It allows for systematic tuning of hyperparameters in models like Lasso and Ridge, leading to better generalization.

5) Elastic Net Regression : Flexibility in Regularization and This model allows for tuning between L1 and L2 .

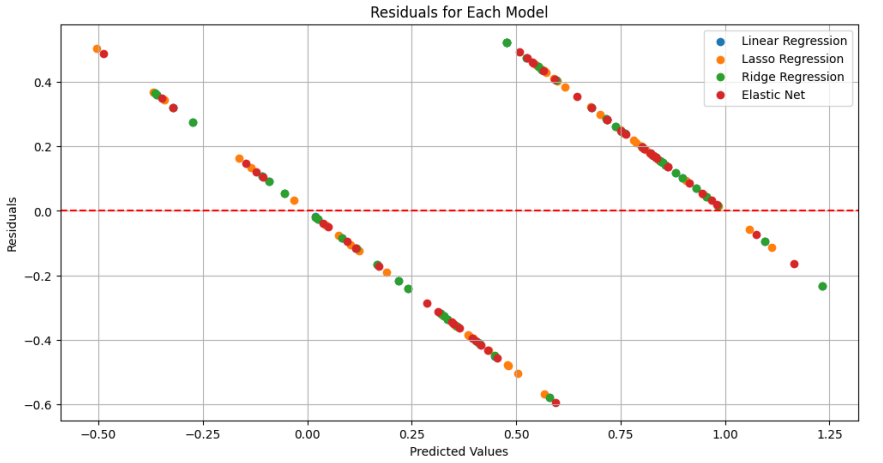
## Exploratory data analysis (EDA) :

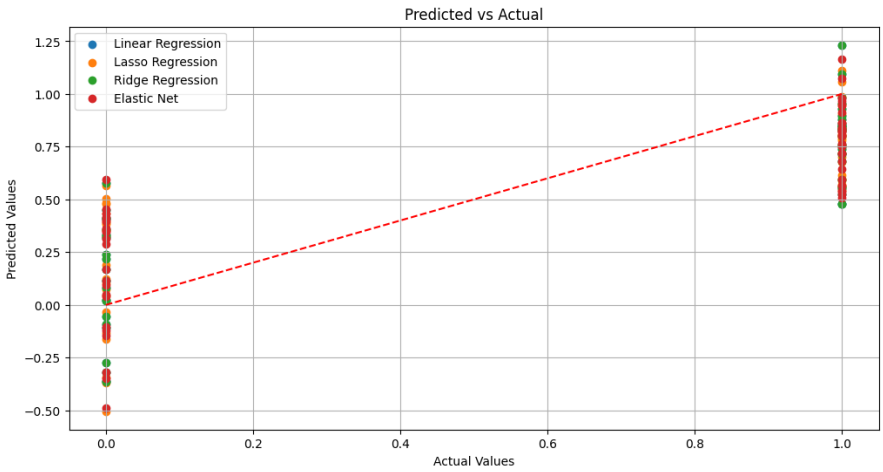

From the correlation Matrix , The goal is to identify whether any two variables are positively or negatively correlated, or if they are independent of each other. In here the dark blue boxes which are highly negative and light color boxes which are high positive to each other. It shows the relation between those variables. Like ,WS and ISI has the most negative relation in here which is -0.72(dark blue) and The diagonal shows the perfect relations in between those variables.




From the box plot we see the fitting for training models. Before it was not regularized ,the variables values are not closely symmetrical. After train and fit the model, it becomes in range and symmetrical.

## Predicted Vs Actual Values for Each Models



Feature scaling improves model performance by ensuring that all features contribute equally, especially in algorithms sensitive to feature magnitude, like Lasso and Ridge regression. Common methods include:

- **Normalization (Min-Max Scaling):** Rescales features to a fixed range, typically [0, 1].
- **Standardization:** Centers features around zero with a standard deviation of one.

The result of the performance of this models :

| Models | Linear | Lasso | Ridge | Cross Validation | ElasticNet |
|---|---|---|---|---|---|
| R2 | 0.54682 | 1.13317 | 0.5642 | 0.6199 | 0.6575 |
| Mean Abs | 0.9847 | 0.94920 | 0.98429 | 0.9820 | 0.9814 |
| Accuracy | 0.938 | 0.9592 | 0.938 | | 0.9796 |
| Precision | 0.9615 | 0.9310 | 0.9615 | | 0.9643 |
| F-1 Score | 0.9434 | 0.9643 | 0.9434 | | 0.9818 |

For the dataset '*Algerian Forest Fires Cleaned Dataset*' Elastic Net Regression model fits better .