Tempest FWI Predictor – A Machine Learning Model to Predict Fire Weather Index

Week 1 & Week 2 Learnings Documentation

Submitted by; Shaik Mohammad Irfan

Date: 20 November 2025

Week 1 & Week 2 Learnings Summary

During the first two weeks of the internship, I focused on building a strong foundation in Linear Regression, Multivariate Regression, and Exploratory Data Analysis (EDA) techniques. These concepts are essential for preprocessing and modelling the Fire Weather Index (FWI) dataset.

1. Understanding Linear Regression

- Linear regression is a supervised machine learning algorithm used to predict a continuous target variable (in our case: FWI).

- It models the relationship as:

$$FWI = \beta_0 + \beta_1 \times Temperature + \beta_2 \times RH + \beta_3 \times Ws + \ldots + \beta_n \times Feature_n + \varepsilon$$

where β_0 = intercept, $\beta_1 \ldots \beta_n$ = coefficients (regressors), ε = error term.

- Goal: Find the best-fitting line by minimizing the Mean Squared Error (MSE) or Sum of Squared Residuals (SSR).

2. Understanding Multivariate (Multiple) Linear Regression

- Extension of simple linear regression to multiple independent variables.

- In our project: FWI is predicted using 9+ features (Temperature, RH, Rain, FFMC, DMC, DC, ISI, BUI, etc.).

- Allows us to capture combined effects (e.g., high Temperature + low RH + high ISI → very high FWI).

- Coefficients (β) show the change in FWI for a one-unit change in a feature, holding all others constant.

3. Systems of Linear Equations

- The Normal Equations form a system of linear equations (one equation per feature + intercept).

- Example with 2 features:

$n\beta_0 + \Sigma x_1\beta_1 + \Sigma x_2\beta_2 = \Sigma y$

$\Sigma x_1\beta_0 + \Sigma x_1^2\beta_1 + \Sigma x_1x_2\beta_2 = \Sigma x_1y$

$\Sigma x_2\beta_0 + \Sigma x_1x_2\beta_1 + \Sigma x_2^2\beta_2 = \Sigma x_2y$

- Solved using matrix inversion or numerical methods (e.g.,numpy.linalg.solve).

- Understanding this helps interpret multicollinearity issues (when X^TX is nearly singular → unstable β).

4. Exploratory Data Analysis Tools Learned

Concept	Purpose	Key Insight from Our FWI Dataset
Histogram	Visualize distribution & skewness of features and target	FWI, Rain, ISI, DMC → right-skewed → consider log transform
Pearson Correlation	Measure linear relationship strength (-1 to +1)	ISI (0.91), DMC (0.88), BUI (0.86) → strongest predictors
Correlation Heatmap	Identify multicollinearity between features	DMC ↔ BUI ↔ DC > 0.9 → high multicollinearity

Key Takeaways Applied to Tempest FWI Predictor

- Performed data cleaning using pandas and linear interpolation (neighbour averaging) → no missing values.

- Generated histograms → identified skewed distributions (especially FWI and Rain).

- Created correlation matrix → selected top features (ISI, DMC, BUI, DC, FFMC, Temperature, RH) for the regression model.