***Data Set Summary:***

***Shivesh Raj Sahu***

***Bu Id:- U28571764***

The dataset provides a comprehensive catalog of solar flares observed by space-based solar observatories. Functioning akin to the unpredictable nature of earthquakes on Earth, solar flares occur on the Sun, their timing hard to anticipate. This collection chronicles the flare activities between 1 May 2010 and 9 October 2017. It meticulously documents the start and end times of 12,455 solar flares, alongside identifying the specific space instruments that recorded each event. This treasure trove of data is a result of an extensive analysis, merging the known orientations of seven distinct instruments with the timings and locations of these solar flare phenomena.

**This code is a comprehensive representation of a typical machine learning project**.

***Here's a breakdown of the code:***

***1. Import Libraries:***

- Necessary Python libraries and tools are imported for data manipulation, visualization, and building machine learning models.

***2. Load and Prepare Data:***

- A dataset containing details about solar flares is loaded using pandas.

- The timestamps are converted from strings to datetime format.

- New columns are created, representing time taken to peak and from peak to end.

- Another feature is engineered which represents the time difference between the current flare and the last one.

- Redundant columns are dropped.

***3. Split the Data:***

- The dataset is divided into independent (`X`) and dependent/target (`y`) variables.

- It is further split into training and testing sets. This helps in training the model on one dataset and validating its performance on another, which is unseen during training.

Results (Visualization):

- A line plot visualizing the "peak to end" times of solar flares against their indices.

- An autocorrelation function plot is used to see how many past days are correlated with the prediction for the current day.

***4. Model Evaluation:***

Each trained model is used to predict on the test data. Mean Squared Error (MSE) and R2 score are calculated to evaluate the performance of each model.

Three regression models (Random Forest, Support Vector Machine, Neural Network) are initialized and trained on the training data.

***6. Results:***

The first 5 rows of the dataset.

The summary of the dataset (data types and non-null counts).

Performance metrics (MSE and R2) for the three regression models on the test data.

My results show:

Random Forest: Has the best R2 score of ~0.28.

SVM: Performs slightly worse with a negative R2, indicating it might be performing no better than a horizontal straight line.

Neural Network: Shows an incredibly high MSE and a massively negative R2.

**The code provides a solid starting point for analyzing solar flare data and predicting the "peak to end" times using various machine learning models.**