**NASA AI: Radiation Exposure Prediction for Astronauts**

**Team Members:**

Maira Tanweer Chachar

Ahad Maredia

**Track Chosen:**

Implementation Track

**Project Overview**

This project aims to build an AI-based solution to predict radiation exposure levels for astronauts using numerical space weather data. Accurate radiation prediction is vital for astronaut safety, particularly for missions beyond Earth's magnetic shield, such as lunar or Martian expeditions.

We explored a mix of regression algorithms and neural network techniques to model radiation exposure based on input features like solar flare intensity and cosmic ray flux.

**Dataset Details**

- Filename: DataSet1.csv  
- Key Features:  
 - Solar flare intensity  
 - Cosmic ray flux  
 - Additional numerical space weather indicators

**Preprocessing Steps:**  
  
**1**. Null Value Removal: Rows containing missing values were removed using df.dropna(). Although the dataset did not have any missing values.  
  
**2.** Outlier Removal: Applied the Interquartile Range (IQR) method to all numerical columns. Any values falling outside 1.5×IQR were removed to reduce skew and improve model accuracy.  
  
**3.** Normalization: All numerical features were normalized using Min-Max Scaling to a [0, 1] range for better model convergence and interpretability.  
  
**4.** Cleaned Output: The final preprocessed dataset was saved as Cleaned\_Data.xlsx.

**Exploratory Data Analysis (EDA)**

- Histograms were created for all numeric features to analyze distribution before and after preprocessing.  
- Correlation Heatmaps helped identify relationships between features, guiding model expectations and feature importance analysis.

**Models Used and Results**

The following machine learning models were implemented and tested to predict radiation levels:

1. Linear Regression   
 Served as a baseline model. While fast and interpretable, it struggled with capturing non-linear interactions in the data.

2. Ridge Regression   
 Added L2 regularization to linear regression. It improved model robustness with minimal overfitting.

3. Lasso Regression   
 Applied L1 regularization, which helped reduce feature redundancy by zeroing out less important features. Performance was similar to Ridge.

4. Bayesian Ridge Regression   
 Introduced a probabilistic approach to ridge regression. It provided useful uncertainty estimates but did not outperform XGBoost or Neural Networks.

5. Random Forest Regressor   
 An ensemble model using multiple decision trees. It successfully modeled non-linear relationships and outperformed linear models on test data.

6. XGBoost Regressor   
 This model gave one of the strongest performances. XGBoost is highly efficient for structured data and captures complex relationships effectively.

7. Neural Network (Keras-based)   
 A custom-built neural network was developed with the following architecture:  
 - Input Layer matching the number of features  
 - Two Hidden Layers with ReLU activations and Dropout for regularization  
 - Output Layer with a single neuron for regression prediction

This deep learning model showed excellent generalization and was comparable to XGBoost in test performance.

**Model Evaluation:**  
  
- All models were evaluated using:  
 - Mean Squared Error (MSE)  
 - R² Score  
- Results were visualized using scatter plots and performance printouts.  
- The best-performing models were XGBoost, Random Forest, and the Neural Network.

**Tools and Environment**

- Platform: Google Colab  
- RAM Requirement: Minimum 8GB  
- Programming Language: Python  
- Libraries Used:  
 - pandas, numpy for data handling  
 - matplotlib, seaborn for visualization  
 - scikit-learn for classic ML models  
 - xgboost for boosted regression  
 - tensorflow.keras for neural networks

**Future Work**

- Use larger or multi-source datasets from NASA to improve model generalization.  
- Build a simple web application or dashboard to make radiation predictions accessible to astronauts, researchers, or mission control.