**Supervised Learning on European Topology 6 Paths Dataset**

**Task Overview**

We have a unique dataset derived from a European topology comprising 6 paths. This dataset includes information on 76 channels with specific features, and the goal is to predict the GSNR (Generalized Signal-to-Noise Ratio), making it a regression task. Your task as an intern will be to apply supervised learning methods to build a predictive model for GSNR and to document your findings and methodologies in a detailed report.

**Dataset Description**

* **Channels**: 76
* **Features**:
  + **Power**: Signal power for each channel
  + **NLI (Non-Linear Interference)**: Measure of interference on each channel
  + **ASE (Amplified Spontaneous Emission)**: Noise figure for each channel
  + **Total Distance**: Total distance the signal travels
  + **Span**: Number of spans in the signal path
  + **Frequency**: Frequency of the channel (not significant for GSNR prediction)

**Label**:

* **GSNR (Generalized Signal-to-Noise Ratio)**: Target variable to predict. Please note that you should select any one GSNR column as the label and discard the rest of them from the dataset.

**Tasks and Expectations**

1. **Data Exploration and Preprocessing**:
   * Analyze the dataset to understand the distribution of features and the target variable.
   * Handle any missing values or anomalies in the data.
   * Perform feature scaling or normalization if necessary.
   * Visualize the relationships between features and the target variable.
2. **Feature Engineering**:
   * Consider creating new features or modifying existing ones to improve model performance.
   * Analyze feature importance to identify which features are most predictive of GSNR.
3. **Feature Selection**:
   * Apply feature selection methods to determine the significance of each feature.
   * Identify and discard the least important features, including the frequency feature, to simplify the model and potentially improve performance.
4. **Model Selection and Training**:
   * Implement various supervised learning algorithms (e.g., Linear Regression, Decision Trees, Random Forest, Gradient Boosting, etc.).
   * Split the data into training and testing sets to evaluate model performance.
   * Use cross-validation techniques to ensure robust evaluation.
5. **Model Evaluation**:
   * Evaluate models using appropriate metrics for regression tasks (e.g., Mean Squared Error, R-squared, Mean Absolute Error).
   * Compare the performance of different models and select the best-performing model.
   * Fine-tune the model parameters to optimize performance.
6. **Reporting**:

* Provide a detailed analysis of the results, including visualizations of model performance and feature importance.

1. **Submission**

Please submit the code and report via email by [21st june]. Make sure your code is well-organized and your report is clear and concise.