**Feature Selection Techniques and Model Performance**

**Project Overview**

This project explores various feature selection techniques and their impact on the performance of a K-Nearest Neighbors (KNN) classification model. We use the Sonar Dataset, which involves predicting whether an object is a mine or a rock based on sonar returns.

**Dataset**

**Dataset Name**: Sonar Dataset  
**Description**: This dataset contains 60 features representing sonar returns at different angles and a binary target variable indicating whether the object is a mine or a rock.

**Project Steps**

**1. Data Preprocessing**

* Loaded the dataset.
* Encoded the target variable.
* Standardized the features using MinMaxScaler.

**2. Initial Model Building**

* Trained a KNN model without feature selection.
* Evaluated the model using accuracy, precision, recall, and F1-score.

**3. Feature Selection Techniques**

* **Filter Method**: Used SelectKBest with the chi-squared statistic.
* **Wrapper Method**: Applied Recursive Feature Elimination (RFE) with Logistic Regression.
* **Embedded Method**: Used Lasso regression for feature selection.

**4. Model Building with Selected Features**

* Trained KNN models using features selected by each method.
* Evaluated the models using the same metrics.

**5. Comparison of Results**

* Compared the performance of the initial model with models using feature selection.

**Results**

| **Model** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| --- | --- | --- | --- | --- |
| Initial Model | 0.88 | 0.96 | 0.85 | 0.90 |
| Filter Method (chi2) | 0.78 | 0.87 | 0.76 | 0.81 |
| Wrapper Method (RFE) | 0.82 | 0.90 | 0.80 | 0.85 |
| Embedded Method (Lasso) | 0.84 | 0.92 | 0.82 | 0.87 |

**Summary of Findings**

* The initial KNN model performed the best without any feature selection.
* The Lasso regression (embedded method) provided the most balanced improvement among the feature selection techniques.
* Feature selection methods reduced the dimensionality but did not significantly improve model performance for this dataset.