1. Forest Cover Type Prediction (forest.ipynb)

- **Problem Statement**: The project aims to predict the forest cover type based on cartographic variables. This is a multi-class classification problem.
- **Dataset Used**
  :
  - The dataset was loaded from covtype.csv.
  - The target variable is 'Cover_Type'.
  - Features include 'Elevation', 'Aspect', 'Slope', 'Horizontal_Distance_To_Hydrology', 'Vertical_Distance_To_Hydrology', 'Horizontal_Distance_To_Roadways', 'Hillshade_9am', 'Hillshade_Noon', 'Hillshade_3pm', 'Horizontal_Distance_To_Fire_Points', and various 'Wilderness_Area' and 'Soil_Type' columns.

- **Methodology and Approach**
  :
  - **Data Preprocessing**
    :
    - Checked for missing values using isnull().sum(); no missing values were found.
    - The features were scaled using StandardScaler.

  - **Model Training**
    :
    - The data was split into training (80%) and testing (20%) sets.
    - Several classification models were trained and evaluated:
      - K-Nearest Neighbors (KNeighborsClassifier)
      - Gaussian Naive Bayes (GaussianNB)
      - Decision Tree Classifier (DecisionTreeClassifier)
      - Random Forest Classifier (RandomForestClassifier)
      - AdaBoost Classifier (AdaBoostClassifier)
      - Gradient Boosting Classifier (GradientBoostingClassifier)
      - XGBoost Classifier (XGBClassifier)

- **Model Evaluation**
  :
  - Models were evaluated based on their accuracy scores on the test set.
  - Confusion matrices and classification reports were also generated for a detailed performance view of some models (specifically Random Forest and XGBoost after hyperparameter tuning).

- **Hyperparameter Tuning**
  :
  - GridSearchCV was used to find the best hyperparameters for Random Forest and XGBoost models.

- **Results and Conclusion**
  :
  - The Random Forest model, after hyperparameter tuning with GridSearchCV (best params: {'max_depth': 20, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 200}), achieved an accuracy of 93.99%.
  - The XGBoost model, after hyperparameter tuning with GridSearchCV (best params: {'learning_rate': 0.1, 'max_depth': 5, 'n_estimators': 200}), achieved an accuracy of 80.09%.
  - The Random Forest classifier was identified as the best-performing model for this dataset among those evaluated.
  - The best Random Forest model was saved to forest_model.pkl.