Earthquake Event Classification Using Machine Learning

Bilal ERÇİN 210717020 Baki TURHAN 210717044

Scope

Classify earthquakes by severity using key attributes like magnitude, depth, and alert levels. Analyze correlations between seismic features to understand impact factors. Support disaster management with actionable insights from data analysis.

Expected Results

A classification model to predict earthquake severity. Insights into factors like depth and magnitude affecting impact.

DATASET

The dataset used for this project originates from the United States Geological Survey (USGS) Earthquake Catalog, accessible at [USGS Earthquake Search]. It contains a comprehensive record of seismic events, structured to enable the classification of earthquakes based on their severity.

Key Characteristics of the Dataset:

- **Number of Rows:** 7,717
- **Number of Columns:** 23
- **Target Variable:** `Alert`
  - Represents the alert level for each earthquake, with four distinct categories:
    - **Green:** Low impact
    - **Yellow:** Moderate impact
    - **Orange:** Significant impact
    - **Red:** High impact

Purpose of the Dataset in the Project

The dataset provides critical information on seismic events, including attributes such as magnitude, depth, location, and alert levels. These features serve as the foundation for:

1. Building a machine learning classification model to predict earthquake severity based on historical data.
2. Analyzing correlations between key seismic parameters (e.g., magnitude, depth) to derive actionable insights into factors influencing earthquake impacts. (Figure 1)

This dataset ensures a data-driven approach to support disaster management efforts, enabling better preparedness and response strategies.

*Also, the dataset used was added to Kaggle accordingly.

https://www.kaggle.com/datasets/29bf2acd173c97ae2af5a6472c40ac8bc1c9c8c640cb959dc095b2cfa1dfa00a

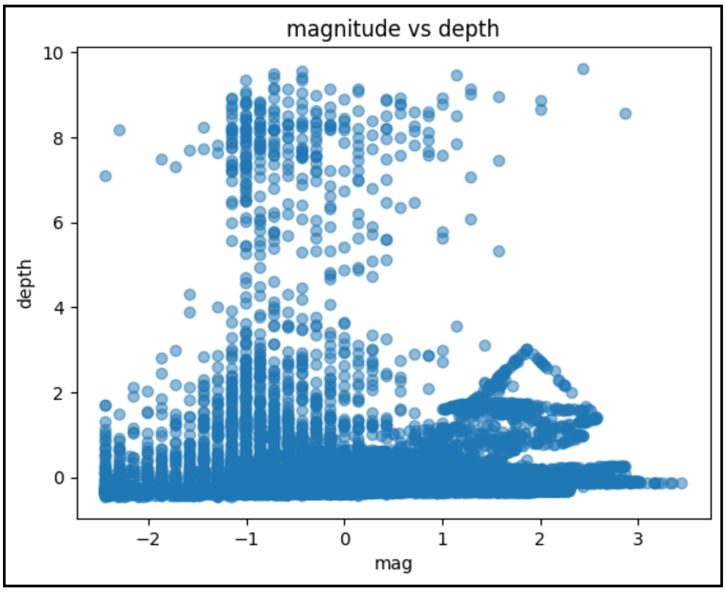

Figure 1

Feature Variables and Feature Engineering

1. Numerical Features
The dataset includes several key numerical variables that describe the characteristics of seismic events:

- **`mag` (Magnitude):** Represents the strength of the earthquake.
- **`depth`:** Denotes the depth at which the earthquake occurred.
- **`latitude` and `longitude`:** Specify the geographical location of the seismic event.

2. Feature Engineering
To enhance the predictive power of the model and better capture the relationships between features, additional variables were created:

- **`impact_score`:** This feature is calculated as the product of magnitude (`mag`) and depth (`depth`). It provides a composite representation of the earthquake's potential impact, combining its strength and origin depth.
- **`log_depth`:** A logarithmic transformation of the `depth` variable was applied to normalize its distribution and reduce the influence of extreme values.

Problem Identified: Class Imbalance in Target Variable
Solution Applied: SMOTE
This approach generates synthetic samples for minority classes, ensuring a more balanced distribution of the target variable and improving the model's ability to predict all classes effectively. (Figure 2 shows the class distributions before and after SMOTE.)

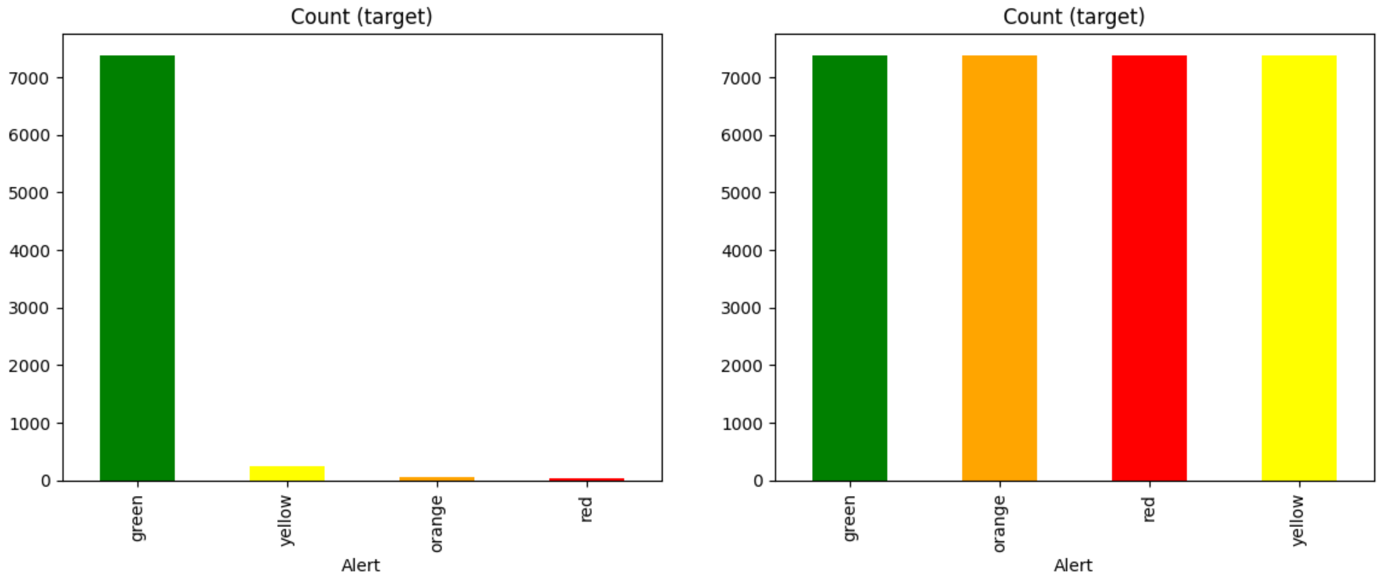

Figure 2

Data Splitting:

- Training Set (80%)
- Testing Set (20%):

Feature Scaling: StandardScaler

To address the varying ranges of features like `depth` and `magnitude`, **StandardScaler** was applied.

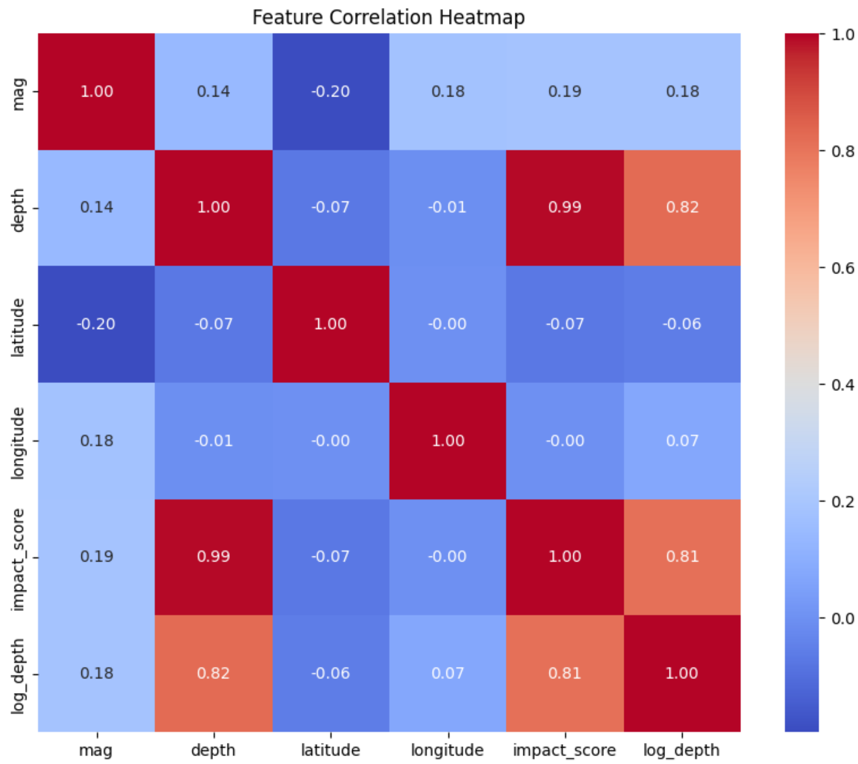

Figure 3

Hyperparameter Tuning and Model Selection

Purpose:
Hyperparameter tuning optimizes model performance by adjusting parameters like tree depth and the number of estimators, ensuring better generalization.

Methodology:
Using **GridSearchCV**, we performed 5-fold cross-validation over a predefined parameter grid for systematic optimization.

Algorithms Applied:

1. **Decision Tree Classifier:** A simple, interpretable model prone to overfitting.
2. **Random Forest Classifier:** An ensemble of decision trees reducing overfitting and variance.
3. **Gradient Boosting Classifier:** An iterative model that captures complex patterns and improves predictions for imbalanced data.

Outcome:
The **Gradient Boosting Classifier** was selected for its superior F1 score, accuracy, and ability to handle class imbalances effectively.(Figure 4)

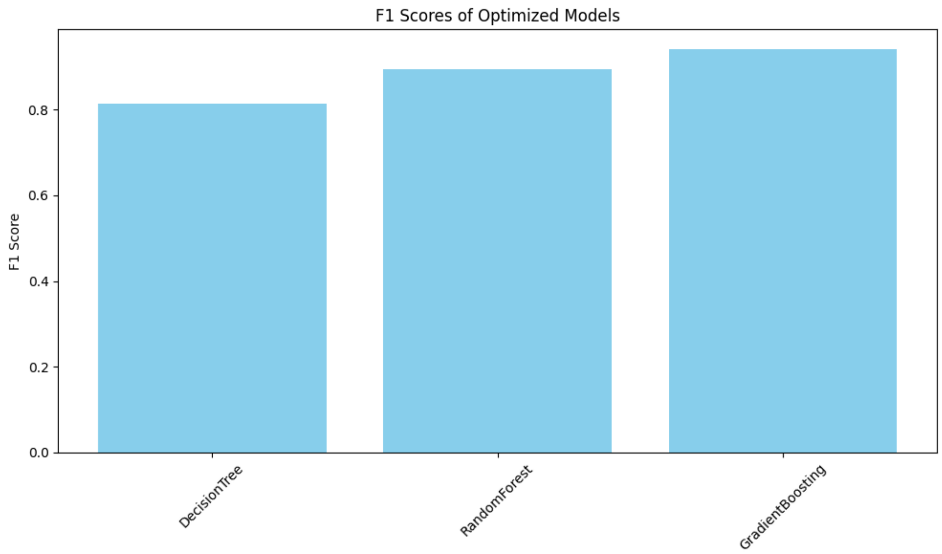

Figure 4

Classification Report for DecisionTree:

	precision	recall	f1-score	support
green	0.95	0.87	0.91	1419
orange	0.85	0.74	0.79	1539
red	0.76	0.85	0.81	1457
yellow	0.73	0.80	0.76	1491
accuracy			0.81	5906
macro avg	0.82	0.82	0.82	5906
weighted avg	0.82	0.81	0.82	5906

Classification Report for RandomForest:

	precision	recall	f1-score	support
green	0.98	0.91	0.95	1419
orange	0.88	0.87	0.87	1539
red	0.89	0.91	0.90	1457
yellow	0.84	0.89	0.86	1491
accuracy			0.90	5906
macro avg	0.90	0.90	0.90	5906
weighted avg	0.90	0.90	0.90	5906

Classification Report for GradientBoosting:

	precision	recall	f1-score	support
green	0.98	0.95	0.96	1419
orange	0.93	0.92	0.93	1539
red	0.93	0.97	0.95	1457
yellow	0.92	0.93	0.92	1491
accuracy			0.94	5906
macro avg	0.94	0.94	0.94	5906
weighted avg	0.94	0.94	0.94	5906

Feature Importance:
Feature importance analysis revealed the most influential variables contributing to earthquake severity predictions, with attributes like `mag` (magnitude), `longitude`, and `latitude` ranking highest. This insight highlights the critical role of these features in determining earthquake impacts.(Figure 5)

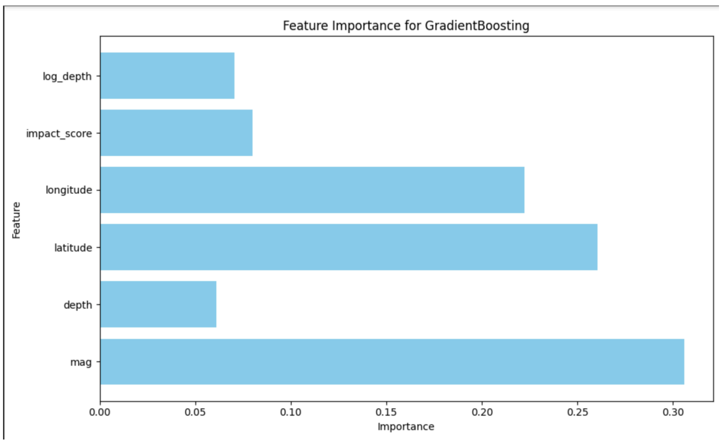

Figure 5

Learning Curve:

Learning curves are used to diagnose **overfitting** or **underfitting** by visualizing model performance on the training and validation sets as the training data size increases.(Figure 6)

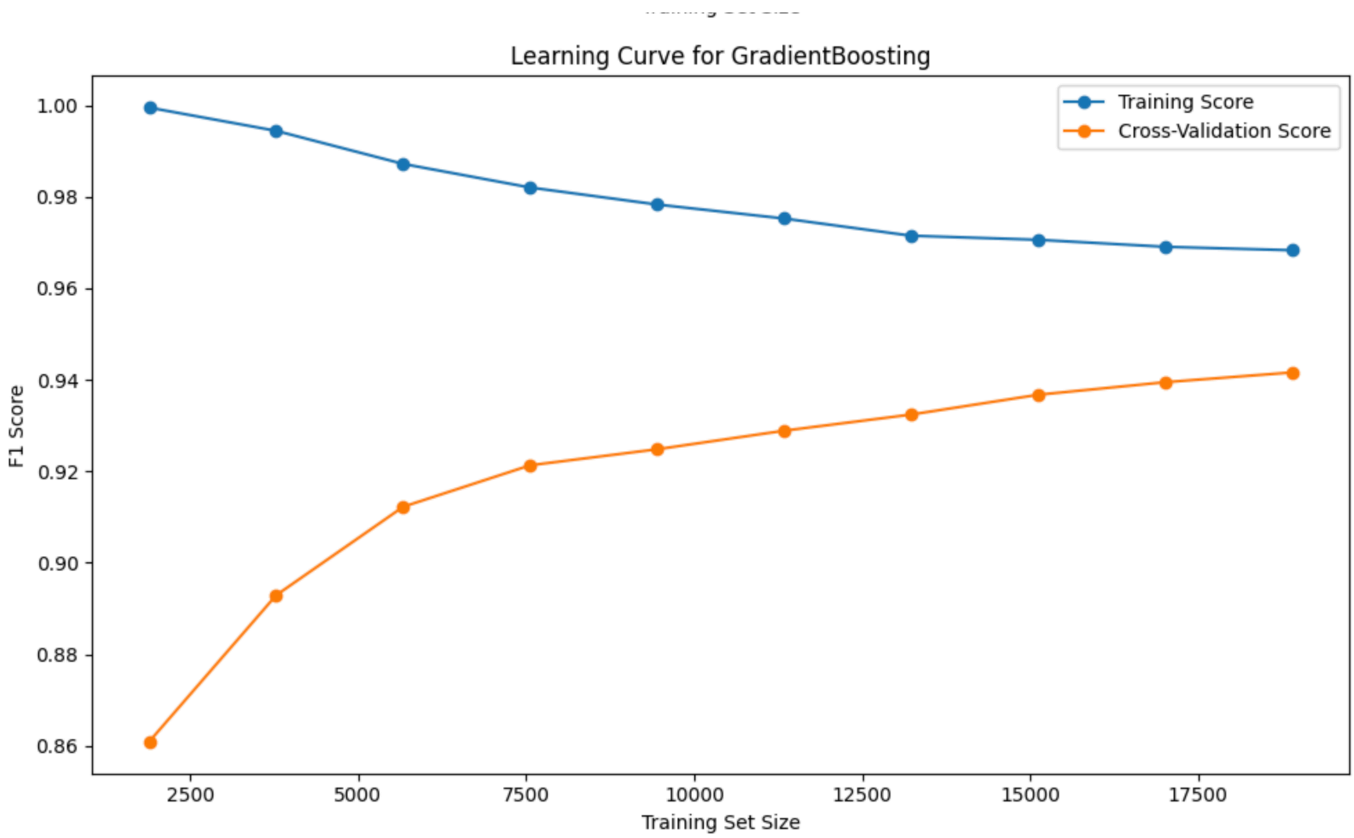

Figure 6

Conclusion:

This project successfully developed a machine learning model to classify earthquake severity, leveraging comprehensive feature engineering, effective handling of class imbalances, and optimized hyperparameters. The selected Gradient Boosting Classifier demonstrated robust performance, providing valuable insights into factors like magnitude and depth that influence earthquake impacts. These results support improved disaster preparedness and response strategies.