Earthquake Occurrence Prediction in Türkiye Using Machine Learning Models: A Rare-Event Classification Approach

MSc Research Project
MSc in Data Analytics

Yunus Emre Hardal
Student ID: x24142590

School of Computing
National College of Ireland

Supervisor: Sheresh Zahoor

National College of Ireland
MSc Project Submission Sheet
School of Computing

Student Name:	Yunus Emre Hardal
Student ID:	x24142590
Programme:	MSc in Data Analytics **Year:** 2025 - 2026
Module:	MSc Research Project
Supervisor:	Sheresh Zahoor
Submission Due Date:	28.01.2026
Project Title:	Earthquake Occurrence Prediction in Türkiye Using Machine Learning Models: A Rare-Event Classification Approach
Word Count:	**7314**
Page Count:	**19**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Yunus Emre Hardal

Date: 28.01.2026

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only

Signature:	
Date:	
Penalty Applied (if applicable):	

Earthquake Occurrence Prediction in Türkiye Using Machine Learning Models: A Rare-Event Classification Approach

Yunus Emre Hardal
x24142590

Abstract

Turkey is among the most active seismic areas in the world and has a continuous occurrence of low-intensity earthquakes along with occasional destructive events. The need for reliable, data-driven predictive tools is apparent. This study investigates two machine-learning methodologies that can be used together to create effective predictive tools, which utilize a large dataset of 493,000 seismic events. One was to classify earthquakes into two categories of destructive and non-destructive and one was to classify earthquakes into 5 magnitude classes (ML, Md, Ml, Mw, MW) using principal components of seismic attributes. Due to the significant imbalance of the destructive event class, the models did not require any's balancing techniques within the training sets because the scientific community understands that the destructive earthquake events are difficult to predict. As a result, despite the fact that all models displayed exceptional overall accuracy, they missed all minority destructive cases, resulting in zero F1 score.

The second task, magnitude type classification, was completed without including the rare event classes to create valid stratified splits for evaluation purposes and for classifying ML, Md, Ml, Mw and MW, respectively. The training data consisted of four numeric parameters (lat/lon/depth/magnitude). Logistic Regression, KNN, Random Forest, Gradient Boosting, AdaBoost and XGBoost were among several models to benchmark. The majority of the classes ML and Md produced excellent predictive capabilities, whereas the remaining classes were difficult to predict. XGBoost provided the highest accuracy score (0.8591). Also, the model improved balance across all classes to some extent.

These findings demonstrate both the power and the limitations of machine learning in applying seismic prediction modelling: while ML can be successfully applied to predicting the magnitude type of an earthquake, applying this methodology to predict a disaster event is much more difficult due to the limited availability of cases and the natural unpredictability of seismic occurrence.

1 Introduction

1.1 Background and Motivation

One of the world's most geodynamically active regions is Turkey which is situated directly over the Alpine-Himalayan seismic belt and is crossed by several major active fault systems including the North Anatolian Fault (NAF), East Anatolian Fault (EAF), and West Anatolian Graben systems. This combination of tectonics makes Turkey seismically active all the time, from many low-magnitude earthquakes, to very large earthquakes that occur infrequently, but can be devastating when they do occur. The damage from these earthquakes also results in a loss of many lives and long-term economic disruption for Turkey.

As more scientists are concerned about a major earthquake coming to the Marmara fault line of North Anatolia, there are growing signs that Istanbul (a city with over 16 million people, which plays an important role in the economy) will be significantly affected during a potential earthquake. Studies have shown that if an earthquake does occur, the result will likely be extremely high numbers of casualties, extensive damage to public and private infrastructure, and continuing negative economic effects on the entire country. Therefore, earthquake prediction, hazard analysis, and the generation of risk-based model data will be needed to support society's preparedness for earthquakes.

Additionally, machine learning can be used to help improve scientific understanding of earthquakes by analyzing very large seismic datasets. Although it is difficult for researchers to predict the exact timing, location, and severity of major earthquakes, researchers will be able to identify patterns in historical seismic recordings that can help improve the understanding of an earthquake (including when, where and how much damage it may do), improve the ability to provide timely warnings to people in the affected area, and create better plans for deploying resources in advance of any earthquake.

1.2 Research Questions

1. Using historical seismic data (e.g., magnitude, depth, latitude, and longitude), can machine learning techniques provide reliable solutions for predicting destructive earthquakes versus non-destructive earthquakes? The purpose of this question is to explore whether it is possible to effectively predict rare but high-impact (i.e., extremely damaging) events by utilizing machine learning techniques on uneven datasets where only a very small percentage of all earthquakes are destructive.
2. How well do various seismic classification methods (ML, Md, Ml, Mw, MW) differentiate between earthquakes by magnitude? Which features within the dataset provide the greatest contribution to predictive capability for machine learning? The purpose of this question is to investigate the extent to which supervised learning techniques are able to accurately model the different scales (or classes) of seismic magnitude when the number of examples within each class is adequate to produce a meaningful result.
3. Of the different types of algorithms available for ML—linear, decision tree ensembles and Gradient boosting approaches—what is the best performance across different

problem types (multi-class, rare-event)? This research will assess and compare the effectiveness of each approach in developing successful ML models for all three tasks.

4. In what ways do the formulation of the imbalance between what is represented by destructive-event prediction versus magnitude-type minority class(es) affect the predictive capabilities of ML models and their evaluation metrics? This question provides insight into the impact of being unbalanced on ML at the model level and at the performance metric level. It offers a better understanding of the limitations imposed on the predictive capability of ML by imbalanced datasets.

Ultimately, this research aims to provide information about the practical application of ML to the analysis of seismic activity in Türkiye—the conditions required, limitations, and strengths of using ML in the seismic area in Türkiye.

1.3 Research Objectives

The primary goal of this research is to explore the potential of supervised machine learning approaches for analysis of Türkiye's seismic dataset(s) as it relates to predicting earthquakes, with the following specific objectives to carry out the broad goal of this research.

1. Preprocess, curate, and construct a comprehensive seismic dataset containing over 493,000 earthquake records. This would entail data cleaning, feature selection, filtering by magnitude type, and creation of a binary destructive/non-destructive label for rare class classification.

2. Build and evaluate supervised machine learning models to assess whether there is a relationship between certain seismic variables (magnitude, depth, latitude, and longitude) and whether or not they represent destructive earthquakes. Additionally, investigate how model performance is impacted by the presence of, and/or extreme class imbalance.

3. To develop a multi-class classification model for determining the magnitude types for earthquakes (ML, Md, Ml, Mw, MW) using training data from the dominant and adequately represented classes; with the exclusion of extremely rare labels to provide an accurate assessment of the models using stratified evaluation.

4. To create benchmark comparisons across a variety of supervised learning algorithms (Logistic Regression, KNN, SGD, Linear SVC, Decision Trees, Random Forests, Gradient Boosting, AdaBoost and XGBoost) to evaluate the predictive power of the models on rare event and size class tasks.

5. To evaluate the performance of each model against the relevant metrics (accuracy, precision, recall, F1-score, confusion matrix and macro averaged) and assess model performance with respect to its application on minority classes.

6. To identify the modelling approaches that demonstrate superior generalisation performance and gain insight into the feasibility, limitations and implications for using machine learning for seismic prediction in Turkey.

2 Related Work

Disaster mitigation includes numerous interdependent elements, with research falling into multiple disciplines related to earthquakes. The areas of study vary from seismology, structural engineering, urban planning and, more recently, data science and artificial intelligence. Within the context of the situation in Türkiye, a considerable amount of research has been conducted on probabilistic seismic hazard assessment, regional risk mapping, and the ability of the built environment to withstand the effects of a strong earthquake (ground shaking).

All existing works show that machine learning (ML) has many potential applications throughout the seismic risk assessment process, from hazard characterisation to building performance, and categorisation of seismic events. However, most previous investigations into ML applied to seismic hazard evaluation either focus on specific minimum thresholds for event magnitude (e.g. ML1.0) or evaluate damage to buildings post-event. In many cases, the sample size for the complete set of observed earthquakes was much smaller than that which would have been anticipated (i.e. an assumed balance of the number of each event type with respect to total) in a typical national-scale catalogue. Consequently, this gap provides the foundation for this study which evaluates ML in both a historical Turkish earthquake catalogue, which has a substantial imbalance due to historic events, and for classifying destructive events on a nationwide scale. As a result, this study aims to demonstrate the experimental limits and potential of ML in a realistic seismological context.

2.1 Earthquake and Seismic Risk in Türkiye

Turkey, located along the Alpine-Himalayan seismic belt, one of the most active seismic and tectonic zones in the world, sees continuous seismic activity because of the interaction of the Anatolian, Eurasian, and Arabian plates. This tectonic setting creates two major fault systems: the North Anatolian Fault (NAF) and the East Anatolian Fault (EAF). Both of these are capable of producing very large and destructive earthquake events. Historical records and contemporary seismic catalogues indicate that Turkey experiences thousands of small-magnitude earthquakes each year. However, although severe earthquake events are infrequent, they have a catastrophic impact when they do occur. According to AFAD (2023), the 6 February 2023 Kahramanmaraş earthquake had a catastrophic impact, including widespread structural collapse across multiple provinces, resulting in over 50,000 deaths and demonstrating the dangerous level of seismic hazard that Turkey possesses.

According to a number of studies, Turkey's seismic risk profile is being analysed through a variety of geophysical and engineering means; this includes the use of seismic hazard maps, fault models, and structural performance data. For example, recent seismic hazard assessments carried out in Turkey show a high likelihood of strong ground shaking in many areas and a large number of vulnerable buildings in the areas studied (Erdik et al., 2009; Kalafat et al., 2021). Studies of the performance of reinforced concrete buildings following the

Kahramanmaraş Quakes show that a combination of design flaws, an ageing building stock, and poor construction practices contributed to the level of destruction caused by these earthquakes (Binici et al., 2023; Yön et al., 2023). Taken together, these studies indicate that to truly understand the earthquake risk in Turkey, consideration needs to be given to not just the majority impact of tectonic forces on the earthquake; but the structural characteristics, socioeconomic status, and urban planning or regulation of each region, in addition to tectonic forces.

The available literature suggests that Turkey's seismic profile consists of frequent low-magnitude seismicity with intermittent short-lived large earthquakes, made worse by a lack of building resilience, as well as ongoing Enhanced Urbanisation. The science of seismology has effectively mapped out a substantial portion of how tectonic movements and the mechanics of Faults work; however, many aspects of Earthquake Prediction remain unsolved due to the uncertainty surrounding when an earthquake will occur. Therefore, interest in using computational methods (for example, Machine Learning) for Earthquake Prediction continues to grow. This larger context defines the current research project, which seeks to assess whether Data-Driven Approaches based on Turkey’s Earthquake Database can assist in obtaining meaningful data to create Models that will aid in Earthquake Prediction.

2.2 Machine Learning for Earthquake Prediction and Seismic Modelling

Machine learning (ML) and artificial intelligence (AI) have become increasingly important in Earthquake Prediction and Seismic Modelling over the past 20 years due to the development of large Seismic Catalogs, as well as advancements in Computational Tools. Initially, work in this area focused on using Artificial Neural Networks (ANNs) to estimate the magnitude, seismic intensity, or probability of occurrence of an earthquake. Researchers like Bodri (2001), Narayanakumar and Raja (2016), and Wang et al. (2009) showed that ANNs were more successful than traditional statistical methods in identifying non-linear relationships between variables in seismic datasets. While ANNs performed well with many training datasets, they often required very large datasets to train and tended to experience problems with overfitting and generalisation, particularly when applied to infrequent and/or extreme earthquakes.

Ensemble techniques such as Random Forests and Gradient Boosting have become popular as they are robust, easy to interpret, and require less data when compared with prior techniques due to their simplicity of use and effectiveness when applied in Earthquake Prediction and Magnitude Estimation tasks (Asim et al., 2020; Mallouhy et al., 2019). However, Random Forests have consistently provided reliable results when used to classify Moderate Earthquake Activity but are unable to predict extremely rare, devastating seismic events. On the other hand, Support Vector Machines (SVMs) have performed well when predicting earthquake magnitude in the mid-range format, but they are subject to the effects of feature scaling and class imbalance on their performance.

The use of imbalanced datasets is a significant obstacle in the context of predicting rare but high-magnitude and/or destructive hazards from a data-driven perspective. In the majority of cases, machine learning algorithms preferentially model majority class instances, which results in high levels of model accuracy but poor recall for the minority class—a fact that has been demonstrated repeatedly in research related to seismic prediction. Despite ongoing experimentation with techniques such as resampling, synthetic data generation, or cost-sensitive training to alleviate this problem, models are unable to generalize and consistently produce reliable results due to the rarity and unpredictability of large earthquake events.

PCA and other dimensionality reduction techniques have been employed to improve the performance of seismic prediction models by minimizing background noise and identifying relevant features from the dataset (Asencio-Cortés et al., 2015; Cilli, 2007), but PCA transformations may inadvertently disrupt domain-related associations between features, which can be critical for seismological interpretation.

In general, literature has shown significant advancement in using ML with seismic data, but it also illustrates that the ability to accurately predict damaging earthquakes is still unsolved because of insufficient data, imbalance of classes, and the inherent complexity associated with seismic processes. Most of the research to date demonstrates good accuracy with respect to classifying or estimating the magnitude associated with earthquakes that are of medium to larger size; however, they tend to be much less successful in classifying or predicting the occurrence of rare events. This is what has led to the current research, which will take on a dual focus of attempting to classify destructive events and predict their magnitude type for earthquakes using Türkiye's national seismic catalogue to evaluate both the strengths and weaknesses of ML models in an imbalanced, realistic context.

3 Research Methodology

To examine the feasibility of using machine learning to forecast earthquakes, this study uses a systematic and structured approach by utilizing Turkey's national earthquake catalog. The methodological framework utilized is based on previously established methodologies presented in the literature but modified to fit the unique characteristics of this dataset, such as the extreme class imbalance associated with the destructive event label and uneven distribution of magnitude type classes. All of these steps were carried out using computational tools based on Python and machine learning libraries that are widely accepted within the industry using a Google Colab environment.

3.1 Data Collection

To perform this research study, an earthquake catalogue containing 493,153 seismic events recorded in Turkey was collected from some of the many available public databases. These records contained many important seismological attributes such as: date and time, latitude, longitude, depth, magnitude, type of magnitude, location description, and identifier for each individual earthquake event. These records cover a long period of time and include many, varying histograms of frequent earthquakes that have small magnitudes, as well as less frequent earthquakes that have large magnitudes and/or destructive value.

3.2 Data Pre-Processing

The Data Preprocessing is an important procedure used to create a clean and useable dataset by eliminating duplicates, correcting any inconsistencies, and providing for a standard format. The actual preprocessing would have consisted of the following:

- Standardisation of the date and time format to a machine-readable structure.
- Removal of entries that contained missing or corrupt data.
- Creation of Binary Destructive Labels (1=Destructive, 0=Not Destructive) since there are only a very small number of Destructive Events compared with all of the Earthquakes at approximately 0.06% of all of the Earthquakes.
- Confirmation of class distribution for both of the classification tasks.

It was determined that no balancing had been required because, due to the extremely low incident rate of Destructive Events, it is widely expected within the scientific community that modelling will be very challenging because of the extreme scarcity of data.

3.3 Data Transformation

The process of data transformation included the transformation of raw attributes into numerical/labelled formats to be understood by algorithms:

- Converting magnitude into categories.
- Normalizing numerical values of continuous features (longitude/latitude/depth/magnitude) to allow for better efficiency in ML algorithms including but not limited to Logistic Regression, K-Nearest Neighbors, Support Vector Machines, Gradient Boosting.

Unlike many previous studies, there is no log transformation that was performed. Linear scaling allowed for superior interpretability. Additionally, log-transformed peaks result in distortion of natural seismic distributions.

3.4 Data Selection

The dataset was divided into several task-specific subsets for the purpose of developing meaningful Machine Learning models:

Classification of Rare Events (i.e. Destructive Classifications)

- The binary outcome variable indicated 492,568 instances of non-destructive and 301 instances of destructive events.
- To model the problem in a realistic manner, neither oversampling nor synthetic data generation were used in this instance.

Classification of Magnitude Types

The distribution of magnitude types for the dataset is disproportionate (as can be seen from the graph below), with the following breakdown:

Type	***Count***
ML	330,159
Md	99,880
Ml	57,352
Mw	3,883
MW	1,594

Four numerical features were selected (Latitude, Longitude, Depth, Magnitude) as the basis for modelling.

3.5 Data Modelling

The Data In terms of Machine Learning, we trained a model (ML) for two different tasks:

Task A (Identifying Destructive vs Non-Destructive)
The following models were evaluated for Task A:

- Logistic Regression
- K-Nearest Neighbour (KNN)
- Stochastic Gradient Descent (SGD Classifier)
- Linear Support Vector Classification (Linear SVC)
- Decision Tree
- Random Forest
- Gradient Boosting
- Adaptive Boost (AdaBoost)
- Extreme Gradient Boosting (XGBoost)

Task B (Classification by Magnitude and Type)
I trained the four classifiers using the features selected for Task B:

- Logistic Regression

- K-Nearest Neighbour (KNN)
- Random Forest
- Gradient Boosting
- Adaptive Boost (AdaBoost)
- Extreme Gradient Boosting (XGBoost)

Ultimately, the XGBoost classifier had the highest accuracy of all the models tested, achieving an accuracy of 0.8591.

All models were developed and trained in Python using scikit-learn and XGBoost libraries using a training/testing data split of 80/20 stratified.

3.6 Data Evaluation

The evaluation methods selected for each of the prediction stages were chosen according to the characteristics of that specific prediction task. The metrics utilized for the popular prediction tasks were as follows:

1. Rare-Event Classification:
- Accuracy
- Precision, Recall, and F1-score
- Confusion matrix
- Recall indicated the most useful characteristics of destructive events because the high percentage of cases classified into the majority class artificially inflated the accuracy measure.

2. Magnitude-type classification:
- Precision, Recall, and F1-score calculated for per-class
- Macro and weighted averages of precision, recall, and F1-score were calculated
- Accuracy calculated on the overall dataset

The models were tested on the robustness of the minority classes (Mw and MW) and the majority classes (ML and Md) were anticipated to perform strongly. The statistical methods listed above provided insight into how the models behaved with respect to actual seismic distributions that are highly imbalanced.

3.7 Predictions and Final Outputs

After training the model:

- Rare events had a consistent 0 recall for destructive events. This confirmed that predicting extremely rare earthquakes is extremely difficult.
- Magnitude type modeling showed that XGBoost performed much better than other models; the cross-class performance of XGBoost is better than any other algorithm used; and recall for XGBoost improved for minority classes.

The final interpretation emphasizes the machine learning capabilities and limitations concerning national scale seismic modeling.

4 Design Specification

This research project has been designed using a modular, layer-based analytical framework. The framework consists of three main components: The Data Persistence Layer, the Analysis and Modelling Layer, and the Evaluation and Interpretation Layer. These components allow for easy reproducibility of the research project methodology, clarity of research methodology, and expansion of the project into the future.

4.1 Data Persistence Layer

The Data Persistence Layer allows for retrieving, storing, and managing the raw seismic data, which contains over 493,000 records of seismic events
Information fields (event time, latitude, longitude, depth, magnitude, magnitude type, and description of the location) are maintained in their original format to retain scientific integrity.

4.2 Analysis and Modelling Layer

The core of the system is the Analysis and Modelling Layer, in which preprocessing and Feature Engineering, model development, and logic of classification are all built in this layer. There are several components that are part of this layer.

Architecture for Classification of Rare Events
The classification of Destructive Events is conceptualised as a supervised binary learning task. Multiple traditional Machine Learning models have been tested including:

- Logistic Regression
- K-Nearest Neighbour (KNN)
- Stochastic Gradient Descent Classifier (SGD)
- Linear SVC
- Decision Tree
- Random Forest
- Gradient Boosting
- Adaptive Boosting
- XGBoost

Each of these models utilised the four fundamental Seismic features: magnitude, depth, latitude, and longitude to train.

Magnitude-type Classification Architecture

The second modelling component consists of multi-class supervised learning to classify an earthquake as one of the five specified magnitude types (e.g., ML, Md, Ml, Mw, and MW), following the removal from the sets of labels that were poorly represented. The following algorithms were benchmarked during model development:

- Logistic Regression
- KNN
- Random Forest
- Gradient Boosting
- AdaBoost
- XGBoost

Classification was carried out using the same four numerical features but resulted in one of five classes as an output. As opposed to destructive classification where class representations are weak (i.e., ML = 330,159 samples; Md = 99,880 samples), in magnitude-type problems class representations are comparatively stable enough for dependable model learning.

4.3 Evaluation Layer

Classification The Evaluation and Interpretation Layer will compare the results of the models in a structured manner for both types of analyses using the metrics identified above.

1. Classifying Rare Events

The evaluation metrics that have been used were:

- Precision
- Recall
- F1 score
- Confusion Matrix
- Macro-Average Metrics

2. Classifying Magnitude Types

With the multi-class classification task, the metrics were assessed both at the individual-class level and across the classes as a whole:

- Individual-class Precision, Recall, and F1
- Macro- and Weighted-Averages
- Overall Accuracy

XGBoost outperformed all of the models on multiple minority classes by being able to incorporate complex, non-linear relationships that exist within seismic data.

4.4 System Integration and Workflow

The analytical workflow for this research follows a linear workflow with modular components:

- Data Retrieval
- Data Preprocessing
- Feature Extraction
- Model Training
- Model Evaluation
- Interpreting Results

5 Implementation

The goal of this project was to create two complementary models using machine learning applications based on the Turkish National Earthquake Catalogue. The first model will classify destructive earthquakes that occur infrequently, and the second will classify earthquakes into different types of magnitude. All the work for this project was conducted using Google Colab, a cloud-based platform, which allowed for the use of many different types and large numbers of datasets for both machine learning projects and its development process.

After cleansing the data, variables that were relevant to the domain were selected to be used in machine learning. The numeric features that were used in both models were Latitude, Longitude, Depth, and Magnitude. The Destructive label was created as a binary indicator and consequently produced a highly unbalanced distribution of 492,568 non-destructive events and only 301 destructive events.

To develop the levels of magnitude that classifies the locations of all earthquake data, the dataset for this part of the analysis required the preprocessing of the magnitude values to a greater degree than the original classifications. The dataset had many different scales for magnitude (ML, Md, Ml, Mw, MW) with certain magnitudes representing less than five events throughout the entirety of the catalogue and could, therefore, not be used when developing the multi-class classification framework via stratified train/test split, thus all rare labels were excluded. The distribution from ML (330,159), Md (99,880), Ml (57,352), Mw (3,883), MW (1,594) made a stable multi-class model possible to develop that could be interpreted accurately. All features included in the models were standardised using the StandardScaler from the Scikit-learn library so as to ensure compatibility with machine learning algorithms that are sensitive to the scale of input features, such as Logistic Regression, Support Vector Machines, and K-Nearest Neighbours.

After preprocessing the data, two pipelines were created for developing and applying the models for rare-class destructive classification purposes.

For rare-class destructive classification, training was conducted using nine different algorithms from the classical categories of exploiting machine learning, including Logistic Regression, K-nearest neighbours, SGD, Linear Support Vector Classify, Decision Trees, Random Forest, Gradient Boosting, AdaBoost, and XGBoost. Model training was conducted on an 80/20 stratification of the data, resulting in a very high accuracy for all algorithms. Since destructive events are so rare, the models were unable to locate any instances, resulting in zero recall and zero F1 Scores for all models.

Classification of Magnitudes by Type (Multi-Class):

The same modelling tool was used for the classification of magnitude types. We evaluated how well the models performed based on their precision/recall/F1 scores separated by class. The modelling approach that used multiple classifiers (XGBoost) and ensemble techniques had better results than linear classifiers, with XGBoost having the best accuracy (0.8591), as well as good balance across minority classes when compared to other algorithms.

6 Evaluation

During this project, many supervised machine learning models were tested with a large dataset of Turkish earthquakes for two complementary purposes: (i) Classification into two Groups: Destructive vs. Non-Destructive (rare prediction) and (ii) Classification of each magnitude of Earthquake (ML, Md, Ml, Mw, MW) using Seismic Attributes. The same underlying Data Set was used in all experiments; also, the Pre-Processing and Design Processes of the Data Set are the same for all experiments.

6.1 Destructive vs. Non-Destructive

Machine Learning was evaluated for its capability with regards to Destructive and Non-Destructive Earthquakes based on four numerical characteristics – Latitude, Longitude, Depth and Magnitude. A Binary Target Variable called "Destructive" was created making it clear that Destructive Earthquake Events are a very small proportion compared to Non-Destructive Events. In fact, of approximately 492,869 cleaned records, only about 301 can be classified as Destructive Earthquakes.

The evaluation results were consistent across all the models. The test accuracy of all models was extremely high (between 0.998 and 0.999). None of the models were able to correctly classify any Destructive Earthquake Events. In every case, the models predicted that the majority of the Events were in the 'Non-Destructive' class and therefore the Recall (Destructive class) was 0.00, Precision (Destructive class) was also 0.00, and the F1-score (Destructive class) was likewise 0.00.

6.2 Magnitude-Type Multi-Class Classification

The second experiment was designed to test a multi-class classification problem in which the class of each event was predicted based on the same four numerical features (magnitude type). Magnitude types to be classified were identified as ML, Md, Ml, Mw, and MW. Classes having fewer than five samples in the dataset were removed to allow proper training and evaluation of the models, as classes with so few samples give insufficient data for generalisation and evaluation stability.

The dataset was divided into a training and a test set with a stratified 80/20 split (training: 394,294 test: 98,574 samples). The features were standardised by applying z-score normalisation.
The experiment results indicate that the ensemble models outperformed the linear models by a considerable margin. Specifically, the results from the experiment indicate that:

- XGBoost achieved the greatest overall accuracy of 0.8591;
- Random Forest performed very closely with an accuracy of approximately 0.859; and
- KNN, and Gradient Boosting (both) achieved accuracies of approximately 0.83 – 0.84.
- The models developed using Logistic Regression, SGD, and Linear SVC produced significantly lower macro-average F1 scores than the other models, and had greatest difficulty classifying the minority class magnitudes.

The development of per-class metrics indicates the following overall trends:

- The two largest classes (majority) ML and Md had high precision and recall in the ML models we tested (usually over 0.80 for each metrics's best performing models).
- The smaller (minority) magnitude classes (Mw and MW) were not as well represented and had a moderate level of precision and recall, but were still much higher than destructive-event prediction tasks.
- The model XGBoost produced the best overall performance based on macro-averages for the F1 scores (approximately 0.67), indicating that lesser represented types of magnitude still benefit from the ability of XGBoost to model non-linear relationships between features in the feature space.

To summarize, although destructive-event prediction is difficult to accomplish with supervised learning methods, when enough magnitude-type classes are present in the training datasets, it is a solvable problem through supervised learning.

6.3 Impact of Class Imbalance on Model Behaviour

The third case study offers an opportunity to compare class imbalance's effect on the two different prediction tasks and see how different families of modelling behave under this condition.

In the binary destructive classification, the class imbalance was severe, with destructive events accounting for less than 0.1% of all samples. Under these circumstances, all the models tested could only arrive at the trivial decision rule of predicting all earthquakes as non-destructive. While this achieves exceptionally high accuracy, it does little to assist in risk mitigation, since there will never be a flag indicating a potentially destructive event.

On the other hand, magnitude-type classifications do face an imbalanced dataset, but the numbers for majority and minority are much more imbalanced than in aspect classifications. In magnitude-type classifications, the numbers of minority observations (Mw, MW) ranged from hundreds (if not more) to several thousand, whereas in aspect classifications these numbers are all in the low hundreds. Fortunately, in the case of magnitude classification, the use of ensemble methods such as Random Forest, Gradient Boosting and XGBoost provided enhanced capabilities to compensate for the impact of class imbalance compared to linear modelling methods. Recall and precision were both considerably better than those of linear models, particularly with regard to the minority classes in this dataset. However, linear models tended to be focused on the majority class and as a result did not capture complex boundary conditions which result in poor macro-averaging performance.

This comparison illustrates an indirect conclusion about methodology and provides an important lesson learned: Not all imbalanced datasets can be handled in the same way; when the size of the minority class is extremely small, classical supervised learning methods will likely not perform as effectively as other methods unless there are additional data (more features describing the minority class) or other formulations of the problem (e.g., anomaly detection) and/or domain-specific characteristics beyond simply magnitude, depth and location.

6.4 Discussion

The evaluation results across all experiments provide a nuanced picture of what machine learning can and cannot realistically achieve in the context of Türkiye's seismic data.

To summarize our two conclusions: All of our models failed to predict Destructive Events within the Rare Event Classification Task because Conventional Supervised Learning cannot produce Operationally Meaningful Predictions For Destructive Earthquakes given Real World Class Proportions along with Limited Feature Sets. This confirms the conclusion made by numerous studies within the literature, that modelling any large scale Rare Event is challenging, due to insufficient representative examples as well as the physically complex nature of Rupture Processes: Even when using very powerful modeling methodologies we find high accuracy in this task is primarily Artifactual & Reflects Model Bias towards the Dominant Non Destructive Class, rather than providing any Predictive Capability.

Our second Conclusion: our machine learning success in Magnitude-Type Classifications demonstrates, while classes of events are represented more accurately within an attribute set & The Prediction Task is Correlated Strongly to Stable, Measurable Attributes: we are able to produce Accurate & Generalizable Results by using Ensemble Based Methods AND This Naturally leads us to The Conclusion that Tree Based & Boosting Models are the More Powerful Type(s) Of Seismological Classification & Regression Methodologies than Linear or Shallow Models, as identified by previous research efforts.

From a critical viewpoint, the current design for the experiments has limitations, as follows:

- There were only four attributes that numerically describe earthquakes and that lack of separate characteristics or similar features like waveforms and regional/geological and fault specific attributes likely limited the ability of the models to distinguish between destructive and non-destructive earthquakes.

- Temporal/sequential modelling (e.g. The time-frame to describe) was not used in the experiments. The lack of temporal modelling may affect the model's capability to accurately discern precursor patterns.
- The fact that the models didn't employ any balancing/cost-sensitive learning was a methodologically correct decision but limited the capability to recall destructive earthquakes.

Potential improvements and further research opportunities that are identified in this study are:

- Inclusion of additional attributes that would include, but are not limited to, focal mechanisms, distance from large fault segments, site characteristics, or ground motion intensity.
- Utilisation of an anomaly detection/one-class classification framework, rather than the traditional supervised method for destructive earthquakes.
- Utilisation of hierarchical/region-specific models, targeting specific tectonic zones (e.g. East Anatolia vs Marmara).
- Evaluation of more recent techniques such as cost-sensitive boosting, focal loss, or, semi-supervised learning for more effective handling of extreme imbalances.

In conclusion, the experiments have shown that while machine learning models can, in general, classify magnitude types, and can serve as an effective analytical tool in the field of seismology, they continue to face significant challenges in their ability to uniformly predict highly destructive earthquakes. The study therefore contributes by clearly illustrating both the **strengths** and **limitations** of current ML approaches when applied to a realistic, national-scale seismic catalogue, and by providing a transparent baseline for future research that seeks to extend or refine this modelling framework.

6.4.1 Handling of Extreme Class Imbalance

A critical methodological decision in this study was the choice not to apply class-balancing techniques for the destructive earthquake classification task. This decision was made deliberately and is grounded in both scientific validity and the nature of the available data. Destructive earthquakes represent an extremely rare class within the national seismic catalogue, accounting for approximately 0.06% of all recorded events. In such cases, artificially altering the class distribution risks introducing patterns that do not reflect the true physical processes governing seismic activity.

Several alternative techniques could have been applied to address the imbalance, including oversampling methods such as SMOTE, random oversampling, class weighting, or cost-sensitive learning approaches. While these techniques are commonly used in machine learning to improve minority-class recall, they rely on the assumption that the minority class is sufficiently represented to allow meaningful generalisation. In the context of destructive earthquakes, where the number of true events is extremely limited, synthetic sample generation would effectively create hypothetical earthquakes that have no physical or geophysical basis. This would undermine the scientific integrity of the analysis and could lead to misleading conclusions regarding predictive capability.

Furthermore, applying balancing techniques in this setting may inflate evaluation metrics such as recall or F1-score without improving real-world usefulness. A model that appears to perform well on artificially balanced data may fail entirely when exposed to real seismic distributions, where destructive events remain exceedingly rare. Since the objective of this research was to assess the feasibility of machine learning under realistic, real-world conditions rather than to optimise metrics in a controlled setting, preserving the true class distribution was considered methodologically appropriate.

As a result, the observed outcome—high overall accuracy combined with zero recall for destructive events—should not be interpreted as a failure of the modelling process, but rather as an empirical demonstration of the limitations of conventional supervised learning when applied to extreme rare-event prediction with limited feature sets. This finding aligns with existing literature and highlights the need for alternative formulations, such as anomaly detection or the inclusion of richer geophysical features, in future research.

7 Conclusion and Future Work

This study investigated the potential of using machine learning techniques to create intuitive models for two separate experimental tasks focused on making earthquake predictions. The two prediction tasks were: (i) distinguishing between the two types of event: destructive and non-destructive; and (ii) classifying earthquakes based upon their magnitudes using their seismic attribute data as the independent variables.

The results of our investigation into the first research question (can earthquakes be predicted analytically by observing their basic seismic attribute data?) indicate that we found none of the supervised learning models exhibited the required capability to differentiate between the two types of event. All models achieved a recall score of zero for destructive events. While the overall accuracy was relatively high due to the overwhelming dominance of non-destructive events, this result confirmed our hypothesis that simple supervised learning models trained on very highly imbalanced data sets with a limited number of distinguishing features will not be able to provide reliable predictive capability for rare destructive earthquakes.

To summarise, the research achieved the objectives of the study regarding the development, implementation and critical evaluation of both modelling pipelines. The study presents machine learning's dual role in seismology, as both a suitable tool for classification of magnitude scales and a limited capability for predicting destructive events, given only basic spatial and magnitude inputs.

The research project faced a number of limitations. The models had access to only four numerical features (latitude, longitude, depth, and magnitude); it did not include more informative features, such as the waveform signature, peak ground acceleration, the regional geology from which the earthquake originated, and indicators of fault-segment.

Meaningful directions for future work include:

- The research includes the integration of seismic waveform data, which may contain precursory patterns or rupture signatures (features not available from catalogue-based

features), as well as region-specific modelling, especially in high-risk areas (e.g., Marmara Region), where tectonic dynamics differ from East Anatolia's region.
- The researchers explored different classification frameworks (e.g., one-class classification frameworks) that may be more appropriate for finding anomalous seismic patterns than for forecasting destructive events.
- Researchers examined external geophysical parameters (e.g., fault proximity, crustal strain rates, changes in coulomb stress, satellite and other methods to measure crustal deformation, etc.) and implemented advanced methodologies to optimise performance (e.g., cost-sensitive strategies, focal loss, and Bayesian uncertainty modelling techniques) for addressing extreme class imbalance.

This research represents a contribution towards predicting future destructive earthquakes, although the prediction of future destructive earthquakes is still an open scientific challenge. However, researchers have provided an evaluation of available machine learning capabilities in a real-world seismic environment and created a framework for future research that integrates better feature sets, incorporates knowledge from a source that is more representative of that area, and examines alternative modelling strategies. The results indicate that although machine learning cannot yet accurately predict destructive earthquake events, machine learning can be used to provide useful analytical information and can be applied in conjunction with broader efforts to assess seismic risk.

References

AFAD. (2023). *6 Şubat 2023 Kahramanmaraş depremleri raporu*. Afet ve Acil Durum Yönetimi Başkanlığı. https://www.afad.gov.tr

Alarifi, A. S., Alarifi, N. S., & Al-Humidan, S. (2012). Earthquakes magnitude prediction using artificial neural network in northern Red Sea area. *Journal of King Saud University – Science, 24*(4), 301–313.

Asencio-Cortés, G., Martínez-Álvarez, F., Morales-Esteban, A., Reyes, J., & Troncoso, A. (2015). Improving earthquake prediction with principal component analysis: Application to Chile. In *Hybrid Artificial Intelligent Systems: 10th International Conference, HAIS 2015* (pp. 393–404). Springer.

Asim, K. M., Moustafa, S. S., Niaz, I. A., Elawadi, E. A., Iqbal, T., & Martínez-Álvarez, F. (2020). Seismicity analysis and machine learning models for short-term low magnitude seismic activity predictions in Cyprus. *Soil Dynamics and Earthquake Engineering, 130*, 105932.

Ayor, D., & Ozkul, B. D. (2023). The strange case of earthquake risk mitigation in Istanbul. *European Planning Studies, 31*(4), 833–854. https://doi.org/10.1080/09654313.2023.2192761

Berhich, A., Belouadha, F.-Z., & Kabbaj, M. I. (2022). A location-dependent earthquake prediction using recurrent neural network algorithms. *Soil Dynamics and Earthquake Engineering, 161*, 107389.

Bikce, M. (2017). Taken and need to take precautions in Turkey's earthquakes. *International Journal of Engineering Research and Development, 9*(2), 24–31.

Binici, B., Yakut, A., Kadaş, K., Demirel, O., Akpınar, U., Canbolat, A., ... & others. (2023). Performance of RC buildings after Kahramanmaraş earthquakes: Lessons toward performance-based design. *Earthquake Engineering and Engineering Vibration, 22*(4), 883–894. https://doi.org/10.1007/s11803-023-2206-8

Bodri, B. (2001). A neural-network model for earthquake occurrence. *Journal of Geodynamics, 32*(3), 289–310.

Cosgun, C. (2023). Machine learning for the prediction of evaluation of existing reinforced concrete structures performance against earthquakes. *Case Studies in Construction Materials, 18*, e02355.

Erdik, M., Durukal, E., Sesetyan, K., Kamer, Y., & Demircioğlu, M. B. (2009). Seismic risk assessment for the city of Izmir, Turkey. *Natural Hazards, 48*, 115–130. https://doi.org/10.1007/s11069-009-9455-3

Ergunay, O. (2007). Türkiye'nin afet profili. *TMMOB Afet Sempozyumu, 5*(7), 1–14.

Holliday, J. R., Chen, C.-C., Tiampo, K. F., Rundle, J. B., Turcotte, D. L., & Donnellan, A. (2007). A RELM earthquake forecast based on pattern informatics. *Seismological Research Letters, 78*(1), 87–93.

Kalafat, D., Kekovalı, K., Kara, M., & Yılmazer, M. (2021). Probabilistic seismic hazard assessment of Muğla, Turkey. *Natural Hazards, 107*, 2017–2039. https://doi.org/10.1007/s11069-021-04633-9

Mallouhy, R., Abou Jaoude, C., Guyeux, C., & Makhoul, A. (2019). Major earthquake event prediction using various machine learning algorithms. In *2019 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)* (pp. 1–7). IEEE.

Niksarlıoğlu, S., & Kulahci, F. (2013). An artificial neural network model for earthquake prediction and relations between environmental parameters and earthquakes. *International Journal of Geological and Environmental Engineering, 7*(2), 87–90.

Ozturk, S., & Alkan, H. (2022). Van Gölü ve civarı (Türkiye) için deprem tahmini ve tehlikesi üzerine istatistiksel bir analiz ve değerlendirme. *Türk Deprem Araştırma Dergisi, 4*(2), 191–209.

Shcherbakov, M. V., Brebels, A., Shcherbakova, N. L., Tyukov, A. P., Janovsky, T. A., & Kamaev, V. A. (2013). A survey of forecast error measures. *World Applied Sciences Journal, 24*(24), 171–176.

Wiemer, S., & Wyss, M. (1994). Seismic quiescence before the Landers (M = 7.5) and Big Bear (M = 6.5) 1992 earthquakes. *Bulletin of the Seismological Society of America, 84*(3), 900–916.

Yön, B., Dedeoğlu, İ. Ö., Yetkin, M., Erkek, H., & Calayır, Y. (2023). Evaluation of the seismic response of reinforced concrete buildings in the light of lessons learned from the February 6, 2023, Kahramanmaraş, Türkiye earthquake sequences. *Earthquake Spectra*. https://doi.org/10.1177/87552930231208120