Anas AlMuslomawi

Layal Farwaji

Suhaib AlKhalidi

**Machine Learning Final Project**

**Earthquake Prediction and Location Classification**

**I. Introduction**

Recent devastating earthquakes in the region of Turkey have begged the question – could anyone have seen them coming? However, seismologists emphasize that accurately forecasting the exact time, location, and magnitude of a major earthquake is currently beyond our scientific capabilities. The unpredictability comes due to the multitude of variables involved in seismic activities. What scientists can do is offer short-term warnings, tens of seconds of notice, once an earthquake has been detected, and there are some advancements that aim to provide even longer warnings, almost a minute or so. Progress is being made in computing risk forecasts for specific regions over the years, and in aftershock forecasting, but unfortunately, predicting earthquakes with absolute certainty ahead of time remains an elusive goal.

The objective of this project was to develop machine learning models for earthquake prediction in Turkey, encompassing both time prediction and location classification. The dataset provided spans earthquake occurrences in Turkey from 1915 to 2023. It includes variables such as date, latitude, longitude, depth, and various magnitude values. Through the training and testing of different models, we aim to get as close as possible to an accurate prediction of the next earthquake to occur.

**II. Methodology**

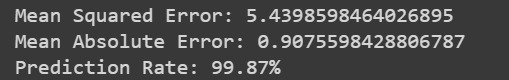
1. Data Preprocessing

The date variable was preprocessed to extract relevant time-related features; year, month, and day. Time of occurrence was also processed. Latitude and longitude data were normalized and the geographic area was segmented into distinct classes using the K-Means clustering technique. The different magnitude value columns were dropped, leaving only the xM (Largest Magnitude) column as a model feature. The newly processed dataset was exported to CSV files that contained only the relevant columns for the time prediction and location classification tasks for subsequent analysis.

2. Time Prediction (Regression)

The dataset was split temporally into training and testing sets, with the training set including all data before 2010, including 2010, and the test set including the rest of the data from recent years. RNN (Recurrent Neural Networks) and LSTM (Long Short-Term Memory) were the models considered for predicting the time of the next earthquake because of their capacity in handling sequential data and potentially capturing patterns from sequences. We observed a challenge in implementing the temporal split in the LSTM model but it still showed positive performance. The Mean Squared Error, Mean Absolute Error, and Prediction Rate performance metrics were employed for model evaluation, and the RNN model had a better performance overall. Since the available dataset is limited, an RNN might generalize better than LSTM with less data. The sequential complexities are too simple for the LSTM model.

RNN evaluation:



LSTM evaluation:



3. Location Classification (Multiclass Classification)

Classification models Random Forest and SVM were implemented to predict the location class of forthcoming earthquakes. They are strong choices for earthquake location classification because Random Forests excel in handling non-linear relationships, and providing robustness to noise, and SVMs also excel in handling non-linear relationships by defining clear spatial boundaries between earthquake location classes. Evaluation metrics such as accuracy, precision, and recall were employed. Accuracy indicates the proportion of correctly classified earthquake locations, precision reflects the model's ability to avoid false positives, and recall denotes the model's capacity to capture all positive instances. Overall, these results suggest that both models are highly effective in classifying earthquake locations, but Random Forest model had a better performance.

Random Forest evaluation:



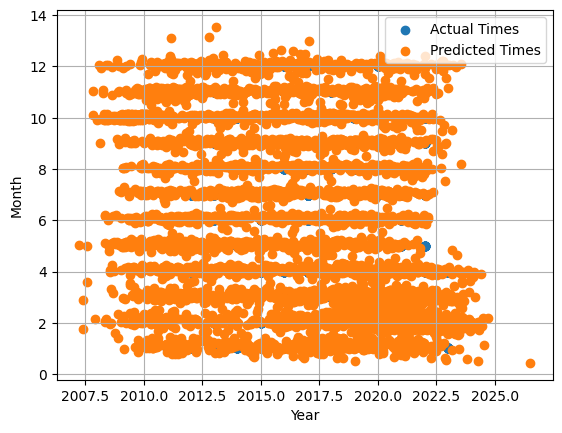
SVM evaluation:



**III. Results**

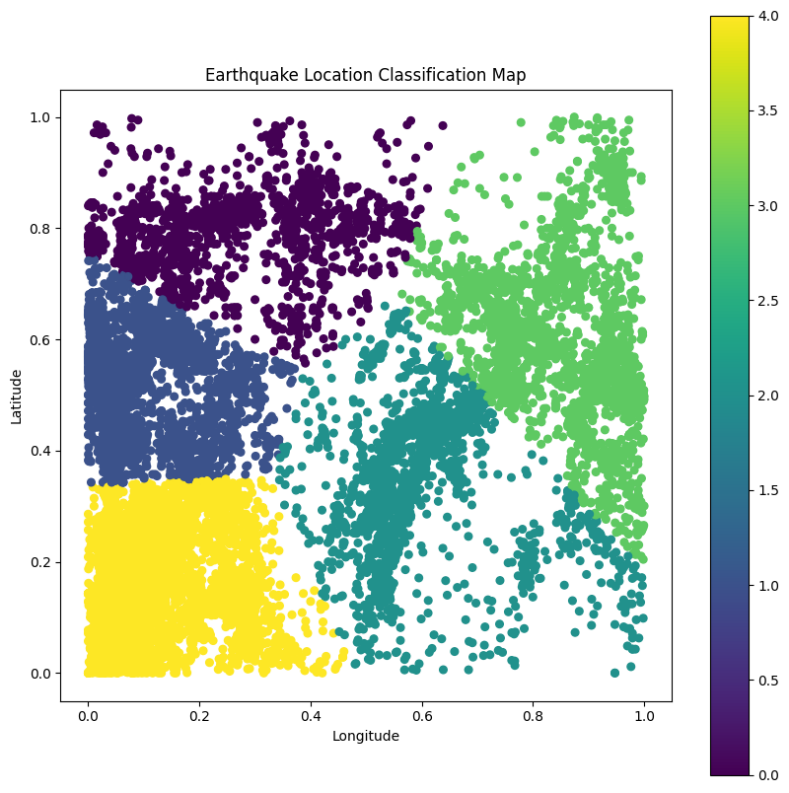
Prediction results were visualized using different techniques.

Time Series Plot (RNN):

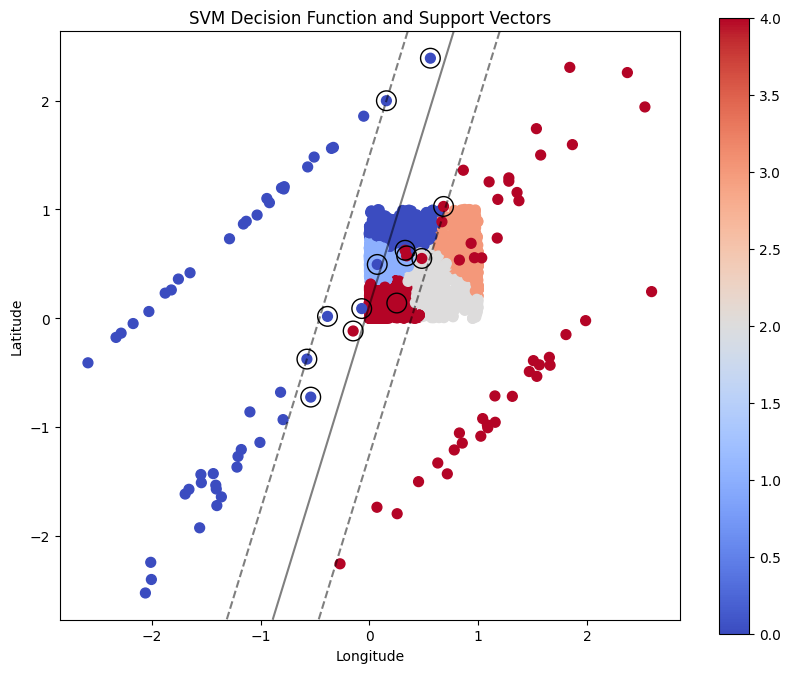


The time series plot shows a correctness of our prediction model as the predicted times cover most of the actual times.

Classification Map (Random Forest):



Classification Map (SVM):



**IV. Conclusion**

Although natural disasters remain mostly impulsive events, the project successfully developed predictive models for earthquake occurrence time and location classification. The chosen models demonstrated promising performance, emphasizing the importance of applying machine learning to earthquake prediction.

**V. References**

<https://www.thenationalnews.com/world/2023/02/08/could-the-turkey-syria-earthquake-have-been-predicted/>

<https://www.sciencedirect.com/science/article/abs/pii/S095741741100827X>

<https://www.javatpoint.com/earthquake-prediction-using-machine-learning#:~:text=Machine%20learning%20has%20the%20ability,a%20complicated%20and%20difficult%20topic>.

<https://practicaldatascience.co.uk/data-science/how-to-reformat-dates-in-pandas>

<https://e-learning.gju.edu.jo/pluginfile.php/215239/mod_resource/content/1/Neural%20Network-1.pdf>

<https://e-learning.gju.edu.jo/pluginfile.php/216862/mod_resource/content/1/SVM%20%5BAutosaved%5D.pdf>