Configuration Manual

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:	**1872**		
Page Count:	**7**		

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):	

Configuration Manual

Yunus Emre Hardal
Student ID: x24142590

1 System Requirements & Environment Setup

All development and experiments for this work have taken place using Google Colab, which provides a Python environment with GPU capabilities. Running the experiments was done using Python 3.10, using the following libraries:

- numpy
- pandas
- scikit-learn
- xgboost
- matplotlib
- seaborn

The dataset was downloaded from the Kaggle public dataset repository, named "Türkiye Deprem Verileri". As no deep learning models were used to create the final classification tasks, all of these experiments could have been run in a normal CPU environment.

The system runs entirely on Google Colab and requires no local installation.

2 Dataset Configuration & Pre-processing Workflow

The earthquake catalogue from the project includes information about over 493,000 large earthquakes, which were cataloged using this data. The following tasks were completed as part of the preprocessing of this dataset:

1. Deletion of invalid rows, or rows that contain missing or duplicated data;

2. Standardizing all date columns to a single uniform time stamp;

3. Filtering the classes for magnitude; • Any class that had less than five occurrences was removed due to the stratified-split constraints and requirements; • The remaining target classes are ML, Md, Ml, Mw and MW.

4. Choosing feature sets;
   • Numerical feature sets will consist of latitude, longitude, depth and magnitude.

5. Train-test division;
   • The dataset was divided into a stratified training/test set ratio of 80/20.

6. No class-balancing techniques (such as SMOTE, random oversampling, or undersampling) were applied for the destructive earthquake classification task. This decision was made deliberately in order to preserve the original statistical distribution of seismic events and avoid introducing synthetic or artificially amplified destructive samples. Given the extremely rare occurrence of destructive earthquakes in the dataset, applying class-balancing techniques could lead to scientifically misleading results by inflating model performance without reflecting real-world seismic behaviour. Therefore, the original class distribution was retained to ensure methodological integrity and realistic evaluation.

3 Model Configuration, Execution, and Reproducibility

Two types of machine learning experiments were performed:

1. Rare-Event Destructive Earthquake Classification

- Models evaluated: Logistic Regression, KNN, SGD, LinearSVC, Decision Tree, Random Forest, Gradient Boosting, AdaBoost, XGBoost
- Target: *Destructive (yes/no)*
- Metrics: Accuracy, Precision, Recall, F1-Score, Confusion Matrix
- Outcome: All models failed to predict destructive cases due to extreme class imbalance. This outcome reflects the dominance of the non-destructive class and highlights a known limitation of standard supervised learning algorithms when applied to highly imbalanced rare-event datasets.

  Execution command:

```
model.fit(X_train, y_train)
pred = model.predict(X_test)
print(classification_report(y_test, pred))
```

2. Magnitude-Type Multi-class Classification

- Target classes: ML, Md, Ml, Mw, MW
- Best model: **XGBoost (Accuracy ≈ 0.8591)**
- Evaluation metrics:
  - Accuracy
  - Per-class Precision/Recall/F1
  - Confusion Matrix

  Sample configuration:

```
xgb = XGBClassifier(
   learning_rate=0.1,
   max_depth=8,
   n_estimators=200,
   subsample=0.8
)
xgb.fit(X_train, y_train)
```

Reproducibility Notes

- All random states were fixed using random_state=42
- Data splits and preprocessing functions are deterministic
- Running the notebook sequentially reproduces identical results

4 File Structure / Project Folder Layout

```
project-root/
│
├── data/
│   └── turkey_earthquakes.csv
│
├── models/
│   └── xgboost_model.pkl
│
├── notebooks/
│   └── experiment.ipynb
│
├── outputs/
│   ├── confusion_matrix.png
│   ├── f1_scores.png
```

5 Execution Instructions

1. Use Google Colab to open the notebook.
2. Set the environment configuration by executing the environment setup cell.
3. Using the Kaggle link provided, retrieve the dataset.
4. Run preprocessing functions.
5. Create the model training loop.
6. Generate performance evaluation plots.
7. (Optional): Save the trained model.

6 Dependencies (pip install list)

- "pip install numpy pandas scikit-learn xgboost seaborn matplotlib"

7 Troubleshooting & Common Errors

Error	Reason	Solution
FileNotFoundError	Kaggle dataset not mounted	Re-run the mount cell
XGBoostError: incompatible version	Local CPU version mismatch	pip install xgboost==1.7
MemoryError	Dataset too large	Use chunk loading or Google Colab Pro

References

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

Aghamohammadi, H., Mesgari, M., Mansourian, A., & Molaei, D. (2013). Seismic human loss estimation for an earthquake disaster using neural network. *International Journal of Environmental Science and Technology, 10*, 931–939.

Al Banna, M. H., Ghosh, T., Al Nahian, M. J., Taher, K. A., Kaiser, M. S., Mahmud, M., Hossain, M. S., & Andersson, K. (2021). Attention-based bi-directional long-short term memory network for earthquake prediction. *IEEE Access, 9*, 56589–56603.

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.

Altun, F. (2018). Afetlerin ekonomik ve sosyal etkileri: Türkiye örneği üzerinden bir değerlendirme. *Sosyal Çalışma Dergisi, 2*(1), 1–15.

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.

Azam, F., Sharif, M., Yasmin, M., & Mohsin, S. (2014). Artificial intelligence based techniques for earthquake prediction: A review. *Science International, 26*(4), 1495–1502.

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. (2021). LSTM-based earthquake prediction: Enhanced time feature and data representation. *International Journal of High Performance Systems Architecture, 10*(1), 1–11.

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.

Berhich, A., Belouadha, F.-Z., & Kabbaj, M. I. (2023). An attention-based LSTM network for large earthquake prediction. *Soil Dynamics and Earthquake Engineering, 165*, 107663.

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.

Buyukkaracigan, N. (2016). Türkiye'de yerel yönetimlerde kriz ve afet yönetim çalışmalarının mevzuat açısından değerlendirilmesi. *Selçuk Üniversitesi Sosyal ve Teknik Araştırmalar Dergisi, 12*, 195–219.

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.

Cui, S., Yin, Y., Wang, D., Li, Z., & Wang, Y. (2021). A stacking-based ensemble learning method for earthquake casualty prediction. *Applied Soft Computing, 101*, 107038.

DeVries, P. M. R., Viégas, F., Wattenberg, M., & Meade, B. J. (2018). Deep learning of aftershock patterns following large earthquakes. *Nature, 560*(7720), 632–634.

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.

Natarajan, Y., Wadhwa, G., Ranganathan, P. A., & Natarajan, K. (2023). Earthquake damage prediction and rapid assessment of building damage using machine learning. In *2023 International Conference for Advancement in Technology (ICONAT)* (pp. 1–5). IEEE.

Nicolis, O., Plaza, F., & Salas, R. (2021). Prediction of intensity and location of seismic events using deep learning. *Spatial Statistics, 42*, 100442.

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.

Pandit, A., & Biswal, K. C. (2019). Prediction of earthquake magnitude using adaptive neuro fuzzy inference system. *Earth Science Informatics, 12*, 513–524.

Sen, S. (2023). Kahramanmaraş depremlerinin ekonomiye etkisi. *Diplomasi ve Strateji Dergisi*, *4*(1), 1–55.

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.

TÜİK. (2022). *Türkiye İstatistik Kurumu bina veritabanı*. Türkiye İstatistik Kurumu. https://data.tuik.gov.tr

Vasti, M., & Dev, A. (2020). Classification and analysis of real-world earthquake data using various machine learning algorithms. In *Advances in Data Sciences, Security and Applications: Proceedings of ICDSSA 2019* (pp. 1–14). Springer.

Wang, Q., Guo, Y., Yu, L., & Li, P. (2017). Earthquake prediction based on spatio-temporal data mining: An LSTM network approach. *IEEE Transactions on Emerging Topics in Computing, 8*(1), 148–158.

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

Demo Artifact : x24142590_MSc_Research_Project_Presentation-Meeting Recording.mp4