Sinopfbd, 2025; 10(2): 690-708 https://doi.org/10.33484/sinopfbd.1751220

Sinop Üniversitesi Fen Bilimleri Dergisi
Sinop Uni J Nat Sci
E-ISSN: 2564-7873

https://dergipark.org.tr/tr/pub/sinopfbd

Worldwide Significant Earthquakes Data Analytics Using Business Intelligence Techniques

Mouhib ALNOUKARİ[1], Anas ABDULAZİZ[1] and Joud ALAKKAD[1]

How to cite: Alnoukari, M., Abdulaziz, A., & Alakkad, J. (2025). Worldwide significant earthquakes data analytics using business intelligence techniques. *Sinop Üniversitesi Fen Bilimleri Dergisi, 10*(2), 690-708. https://doi.org/10.33484/sinopfbd.1751220

Case Report

Corresponding Author
Mouhib ALNOUKARİ
mouhib.alnoukari@gmail.com

ORCID of the Authors
M.A: 0000-0002-3982-2074
A.B: 0009-0005-2632-629X
J.A: 0009-0001-3499-220X

Received: 26.07.2025
Accepted: 27.11.2025

Abstract

Earthquakes are among the most upsetting natural events, capable of causing enormous destruction and loss of human lives. Understanding and mitigating their impacts requires sophisticated analysis tools. One viable strategy entails applying business intelligence (BI), which can effectively examine extensive historical earthquake records alongside with real-time seismic data. By exploring patterns hidden within this information, researchers can potentially uncover early signals indicating an impending earthquake. In this paper, the aim is to apply BI techniques to comprehensivly analyse earthquake data collected between 1843 and April 2025, focusing on significant seismic events with magnitudes greater than five on the Richter scale. To achieve this, a comprehensive seismic data warehouse was established to enable systematic analysis of past earthquake occurrences. Through the request for data mining methods, this research investigates the complex relationships between geological features and tectonic movements. These methods include clustering techniques that group similar earthquakes to help classify potential precursors to larger seismic events. Additionally, classification approaches categorize earthquakes by their severity low-risk, medium-risk, and high-risk while regression analysis forecasts specific earthquake features such as magnitude, depth, and geographical location. Moreover, the research employes time-series analytical method to forecast future occurrences of significant earthquakes worldwide. Such predictive insights are vital not only for understanding seismic activity patterns but also for improving existing forecasting systems. PowerBI is used for data analytics and visualization to facilitate the analysis of global large-scale earthquake data, reflecting their evolution, impact, and recurring locations; data from 1843 to April 2025 were compiled and examined. Ultimaly, the goal of this work is to refine and advance earthquake analysis tools, thereby enabling more accurate predictions and enhanced preparedness strategies worldwide.

Keywords: Significant earthquakes, business intelligence, data mining, seismic data analytics

İş Zekâsı (BI) Teknikleri Kullanılarak Dünya Çapındaki Önemli Deprem Verilerinin Analizi

[1]Syrian Private University, Faculty of Engineering, Department of Informatics Engineering, Damascus, Syria

Öz

Depremler, büyük yıkıma ve can kayıplarına yol açabilen, en sarsıcı doğal olaylar arasında yer alır. Bu etkileri anlamak ve azaltmak, gelişmiş analiz

araçlarının kullanılmasını gerektirir. Uygulanabilir stratejilerden biri, kapsamlı tarihsel deprem kayıtları ile gerçek zamanlı sismik verileri birlikte inceleyebilen iş zekâsı (BI) yöntemlerinin kullanılmasıdır. Bu verilerin sunduğu gizli kalıpların incelenmesiyle, araştırmacılar yaklaşmakta olan depremlere işaret edebilecek erken sinyalleri potansiyel olarak ortaya çıkarabilir. Bu çalışmada amaç, 1843 ile Nisan 2025 yılları arasında kaydedilen ve Richter ölçeğinde beşin üzerindeki büyük sismik olaylara odaklanan deprem verilerini kapsamlı biçimde analiz etmek için BI tekniklerini uygulamaktır. Bu hedef doğrultusunda, geçmiş deprem olaylarının sistematik bir şekilde incelenmesini mümkün kılan kapsamlı bir sismik veri tabanı oluşturulmuştur. Veri madenciliği yöntemlerinden yararlanılarak jeolojik özellikler ile tektonik hareketler arasındaki karmaşık ilişkiler araştırılmıştır. Bu yöntemler arasında, benzer depremleri gruplandırarak daha büyük sismik olayların potansiyel öncüllerini sınıflandırmaya yardımcı olan kümeleme teknikleri yer almaktadır. Ayrıca, sınıflandırma yaklaşımları depremleri düşük riskli, orta riskli ve yüksek riskli olarak kategorize ederken; regresyon analizleri büyüklük, derinlik ve coğrafi konum gibi belirli deprem özelliklerini tahmin etmektedir. Bunun yanı sıra, zaman serisi analizi yöntemi dünya genelinde gelecekte meydana gelebilecek önemli depremlerin öngörülmesinde kullanılmıştır. Bu tür öngörüler, yalnızca sismik aktivite kalıplarının anlaşılması için değil, mevcut tahmin sistemlerinin iyileştirilmesi açısından da büyük önem taşımaktadır. Küresel ölçekte büyük deprem verilerinin analizi, gelişimi, etkisi ve tekrar eden konumlarını yansıtmak amacıyla Power BI kullanılmış; 1843'ten Nisan 2025'e kadar olan veriler derlenerek incelenmiştir. Sonuç olarak, bu çalışmanın amacı deprem analiz araçlarını geliştirmek ve iyileştirmek; böylece daha doğru tahminler yapılmasını ve dünya genelinde hazırlık stratejilerinin güçlendirilmesini sağlamaktır.

This work is licensed under a Creative Commons Attribution 4.0 International License

Anahtar Kelimeler: Önemli depremler, iş zekası, veri madenciliği, sismik veri analizi

Introduction

Earthquakes are powerful natural events that pose serious threats to human life and property. Because of their complexity, they remain difficult to fully understand, and to this day no single scientific theory can explain all seismic activity. Additionally, there is still no universally reliable method for predicting earthquakes. Nevertheless, recent years have seen meaningful progress in identifying contributing factors and refining predictive methods [1]. While predicting earthquakes remains a difficult challenge, it is crucial for advancing early warning systems and disaster-mitigation strategies [2]. Traditionally, prediction efforts have focused on analyzing seismic data, including earthquake catalogs and seismic waves. Statistical methods such as linear regression and clustering have been used to detect patterns and correlations between seismic events and underlying geological or geophysical variables. However, these methods have shown limited effectiveness, mainly due to their lack of precision and their inability to provide timely, actionable alerts [2]. The use of satellite data offers a promising addition to the field. Satellites can collect a wide range of crucial information, including measurements of surface deformation, soil moisture, and vegetation cover. These variables can be used to detect patterns that may signal increased seismic risk. For example, changes in surface deformation can indicate stress accumulation in the Earth's crust, which may serve as an early-warning indicator [3]. Artificial

intelligence (AI), particularly machine learning (ML), has become a key component in earthquake prediction research [4]. ML models trained on large historical datasets can identify complex patterns that might help forecast future seismic events. Deep learning techniques, which use artificial neural networks, have demonstrated potential in detecting early signs and estimating the characteristics of earthquakes. Moreover, natural language processing (NLP) can analyze text sources—such as news reports and scientific literature—to extract useful insights about earthquake trends that may not be visible in seismic data alone [4]. ML stands out for its ability to process large volumes of data and reveal relationships that might otherwise go unobserved [1-4]. These algorithms are particularly effective in modeling nonlinear behavior, which is common in seismic activity. Additionally, ML models can be deployed in real-time applications, making them suitable for integration into modern early warning systems [4]. They have also been used to estimate potential earthquake damage by analyzing past incidents and identifying damage-related patterns [5]. Several ML techniques serve this purpose. In supervised learning, algorithms are trained with labeled data that link input features to known damage levels, helping predict the likely impact on buildings and infrastructure. Unsupervised learning, meanwhile, uncovers patterns in unlabeled data that can also indicate potential risks. Ensemble learning methods combine the strengths of multiple models to improve prediction accuracy [5]. These algorithms can also process a wide range of inputs, from geological measurements to satellite imagery and ground deformation records [3]. Their strength lies in their ability in filtering vast datasets to detect subtle indicators of future seismic activity. Commonly used ML models include artificial neural networks (ANNs), support vector machines (SVMs), and random forests (RFs), all of which have been applied to predict earthquake parameters, including timing, magnitude, and epicenter [6]. Data mining is central to this workflow, forming a key stage within the broader field of knowledge discovery: It involves aggregating data from multiple sources, preparing it, and applying analytical algorithms within a structured data-warehouse environment [7]. Data mining methods have shown considerable promise in seismic forecasting, particularly due to their ability to handle extensive datasets and reveal meaningful patterns [8]. Such patterns might include changes in seismic activity, ground movement, or surface fluid dynamics. They can also detect precursors minor changes in the Earth's crust that may signal the onset of an earthquake [8]. Several types of data mining techniques have been deployed for earthquake forecasting [8]. Clustering methods are used to identify groups of similar seismic events, potentially indicating future activity. Classification methods can categorize earthquakes by their predictable risk levels (e.g., low, medium, high), while regression models estimate key features such as magnitude, depth, or location. Neural networks, in particular, have gained popularity in this field. For instance, [9] proposed a graph neural network (GNN)-based method to analyze the spatial and temporal structure of seismic events. Applied to the North Anatolian Fault, this technique helped uncover recurring patterns such as clusters and propagation paths, providing valuable insights into how earthquakes unfold in complex fault systems. Similarly, [10] applied a convolutional neural network (CNN) to the Parkfield

earthquake sequence, classifying temporal patterns that could aid future forecasts. Other researchers have explored hybrid approaches. [11] for example combined long short-term memory networks (LSTM) with support vector machines (SVMs) to detect early warning signals. LSTMs were used to capture long-term trends in seismic data, while SVMs categorized observations into seismic or non-seismic groups. [12] advanced the field by applying a multilayer perceptron (MLP), a type of neural network, to historical seismic records. Their model proved highly accurate, achieving up to 80% accuracy in predicting earthquake events. This performance was comparable to the outcomes reported by [13] and [14]. In contrast, alternative methods such as SVMs and curve-fitting models achieved lower accuracy rates of around 70% and 65%, respectively [12]. [12] also noted that the MLP model offered longer response times up to 24 hours compared with traditional systems, which often provide only brief lead times. This study highlights the integration of machine learning, data mining, and business intelligence tools for analyzing global earthquake data. It describes the development of a seismic data warehouse and a supporting BI system based on digital records of significant earthquakes (magnitude greater than 5) from 1843 to April 2025. By analyzing this information, the research aims to classify seismic events by magnitude, depth, and location, and to apply predictive models to forecast earthquake activity in high-risk zones.

The key objectives of this study are:

- To construct a dedicated data warehouse tailored to global seismic events.
- To classify major earthquakes by magnitude and depth.
- To estimate the frequency of significant future earthquakes in various regions using data mining algorithms applied to the historical earthquake data.
- To use the PowerBI tool for data analytics and visualization to facilitate the analysis of global large-scale earthquake data.

Preliminary information

Business intelligence is "the use of all organization's resources: data, applications, people and processes in order to increase its knowledge, implement and achieve its strategy, and adapt to the environment's dynamism" [7]. BI serves as a pivotal tool across various sectors, enhancing decision-making and operational efficiency. Its applications span from data collection and analysis to real-time reporting, enabling organizations to leverage insights for strategic advantage [15, 16]. The use of BI and data analytics in the context of significant earthquakes involves applying data visualization and analytical frameworks to improve decision-making and recovery processes. This approach is especially pertinent in regions susceptible to seismic activity, such as Indonesia, where BI tools like Tableau are used to visualize earthquake data, including location, depth, and magnitude, to support decision-making and preparedness strategies [17, 18]. Furthermore, modeling frameworks are developed to understand business recovery times after an earthquake, which is essential for reducing business interruptions and

economic losses [19]. These methodologies highlight the potential of BI in converting raw data into actionable insights for disaster management and recovery. While the focus on earthquake data analytics using BI tools is promising, it is essential to consider challenges related to data accuracy and integration across different platforms. The effectiveness of BI in disaster management depends on the quality and comprehensiveness of the data, as well as the ability to adapt frameworks to diverse contexts and industries [17]. Data mining techniques, including clustering analysis, have been used to uncover patterns in earthquake data. For example, clustering algorithms classified seismic events based on their magnitude and location, facilitating the identification of high-risk zones [20, 21]. These methods have also been applied to forecast future earthquake magnitude by analyzing historical data from sources such as the United States Geological Survey (USGS) [22]. Data-driven approaches have been central to seismic risk assessment and hazard forecasting. For instance, ensemble Kalman filters and particle flow filters have been used to estimate stress and velocity variables, improving the accuracy of probabilistic forecasts [23]. These methods have been tested on models such as the Burridge-Knopoff 1D model, demonstrating their robustness in chaotic scenarios [23]. Data warehousing plays a vital role in organizing and storing seismic data for efficient analysis. A study focused on designing a data warehouse to assess earthquake risk in healthcare facilities highlights the significance of such systems for critical infrastructure [24]. The proposed model employs a fact constellation schema and incorporates public legislation to identify at-risk buildings. This approach ensures that healthcare facilities essential during and after earthquakes can be effectively evaluated for earthquake risk. Another study emphasizes integrating stream processing frameworks with traditional data warehousing to enhance real-time analytics [25]. This hybrid approach optimizes data flow, maintains consistency, and reduces latency, making it ideal for real-time earthquake monitoring and analysis. Data analytics and data warehousing have revolutionized the field of seismology, allowing researchers to process vast amounts of seismic data, detect patterns, and make informed decisions. The incorporation of machine learning, real-time stream processing, and geo-visual analytics has opened new opportunities for earthquake prediction, risk assessment, and disaster management. As technology continues to advance, the potential for further developments in this field is substantial. Integrating machine learning algorithms with BI tools has transformed earthquake prediction and analysis. Techniques such as Long Short-Term Memory (LSTM) networks have been used to forecast earthquakes by leveraging spatio-temporal relationships in seismic data [11, 18]. Furthermore, hybrid models that combine classification and regression algorithms have been developed to predict seismic intensity distributions, surpassing traditional Ground Motion Prediction Equations (GMPEs) [26]. Real-time data integration and analysis are essential for earthquake early warning systems. Machine learning techniques, such as Random Forest and Gradient Boosting, have been trained on datasets from regions like Southern California and India to enhance the accuracy and robustness of these systems [27, 28]. These systems not only detect seismic signals but also identify phases, thereby improving the reliability of early warnings [27]. Several machine learning algorithms

have been evaluated for their effectiveness in predicting earthquake magnitudes. Random Forest and Gradient Boosting models have consistently outperformed other algorithms, achieving high R-squared scores and demonstrating their suitability for seismic data analysis [29]. Furthermore, deep learning techniques such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have been applied to vision-based tasks and sequential data modeling, respectively [30]. Despite progress in BI and machine learning techniques, several challenges persist. These include the non-Gaussian distribution of state variables during seismic cycles, the complexity of spatio-temporal data, and the need for novel algorithms specifically designed for earthquake prediction [23-31]. Nevertheless, the growing interest in this field, as shown by bibliometric reviews, underscores the potential for further innovation and collaboration [31]. The field of seismology has benefited significantly from the advent of big data. A review on Big Data Seismology highlights the opportunities and challenges presented by large, data-dense inquiries [32]. The development of new sensor systems, improvements in computing power, and advanced algorithms have enabled researchers to process vast amounts of seismic data, leading to new insights into earthquake processes and Earth structure. Big Data Analytics in Seismology enhances the understanding of earthquake processes and 4D Earth structure by utilizing large, data-rich datasets. It leverages new sensor systems, improved computing, and advanced algorithms to analyze extensive seismic data, leading to significant scientific progress [32]. Big Data Analytics can leverage advanced technologies such as Distributed Acoustic Sensing and dense nodal arrays to extract valuable insights from complex seismic data [33]. Additionally, Big Data Analytics is applied for rapid fracture identification within extensive seismic datasets, using spatio-temporal amplitude statistics and machine learning techniques to automate feature interpretation, thereby improving the efficiency and accuracy of subsurface feature analysis [34]. The analysis of earthquake data employs a range of techniques and tools, each with distinct strengths and applications. This overview compares several notable methods, including Tableau Public, LSTM Networks, Random Forest, clustering algorithms, HLO-RF model, and particle flow filters, highlighting their contributions to enhancing earthquake prediction and monitoring. Tableau Public is primarily used to visualize seismic data, allowing researchers to build interactive dashboards that facilitate exploration of complex datasets [17, 18]. Long Short-Term Memory (LSTM) networks are especially effective for time-series forecasting, including spatio-temporal forecasting, because they effectively capture temporal dependencies in seismic data [18]. Random Forest can assess the importance of different seismic features, helping identify key predictors of earthquakes [28, 29]. Clustering algorithms such as K-Means and DBSCAN are applied to classify seismic events and reveal underlying patterns and anomalies [20]. DBSCAN is particularly adept at detecting unusual seismic activities, which can be critical for early warning systems [21]. The HLO-RF model combines hierarchical learning with Random Forest, enhancing predictive performance by leveraging multi-level data structures [39]. This model aims to improve predictions by considering both local and global seismic patterns [39]. Particle Flow Filters are used for real-time data assimilation,

providing updated forecasts as new seismic data arrive; they are especially useful when seismic conditions change rapidly, enabling adaptive forecasting [23]. Table 1 presents a comparative analysis of the tools and techniques used in earthquake data analytics.

Table 1. *Comparative analysis of techniques and tools used for earthquake data analytics*

Technique/Tool	Application	Citation
Tableau Public	Visualization of earthquake data (location, depth, magnitude)	[17, 18]
LSTM Networks	Spatio-temporal forecasting of earthquakes	[18]
Random Forest	Earthquake magnitude prediction and early warning systems	[28, 29]
Clustering Algorithms	Identification of high-risk zones and seismic event classification	[20, 21]
HLO-RF Model	Real-time analysis of geo-social data for earthquake monitoring	[39]
Particle Flow Filters	Probabilistic forecasting and stress estimation in seismic cycles	[23]

A study on seismic big data handling in the oil and gas industry proposes a unified technique using Hadoop and Hive to process large datasets [35]. This approach reduces latency and improves time efficiency, making it suitable for real-time applications. Table 2 summarizes the key data analytics approaches for earthquake research.

Table 2. *Key data analytics approaches for earthquake research*

Approach	Description	Citation
Data Warehousing	Organizes and stores seismic data for efficient analysis and risk assessment.	[24, 36]
Real-Time Stream Processing	Enables real-time analytics and visualization of seismic data.	[25]
Machine Learning	Predicts earthquake magnitude and categorizes seismic events.	[29, 40-43, 45]
Big Data Analytics	Processes large datasets to uncover new insights into earthquake processes.	[32, 35, 39]
Geovisual Analytics	Visualizes spatio-temporal data for pattern identification.	[46]
Deep Learning	Improves accuracy in seismic event detection and forecasting.	[37, 38, 47]
Cloud Computing	Facilitates large-scale processing of seismic data.	[48]
High-Quality Datasets	Provides ground truth for training machine learning models.	[49]

Worldwide Significant Earthquakes Data Analytics Results and Discussion

Analyzing earthquake data using BI tools has not attracted much interest from researchers. Few studies have addressed this area. [18] examined earthquake records from 1900 to 2024, focusing on both temporal and spatial aspects. This analysis provided insights into the frequency, intensity, and impacts

of seismic events over time, yielding valuable information about their patterns and behaviors. The research obtained earthquake data from two primary sources: Wikipedia and the United States Geological Survey (USGS) API. This comprehensive dataset spans from 1900 to 2024, enabling thorough analysis of seismic occurrences over time. The use of Tableau for data analytics and visualization is emphasized as a key contribution. This approach enhances the clarity and accessibility of the findings, making complex data more comprehensible for stakeholders and the general public [18]. [50] examined earthquake records from an Indonesian agency. They developed an interactive visualization of earthquake history by combining spatial and temporal data. This work uses information from the BMKG (Indonesian Meteorology, Climatology, and Geophysics Agency). The primary dataset comprises historical earthquake phenomena in Indonesia from 2004 to 2019. It contains vital details about past seismic occurrences, which are essential for understanding patterns and trends in earthquake activity. By leveraging BMKG data, the study integrates multiple data variables, including occurrence time, geographic coordinates, depth, and magnitude of earthquakes. This thorough approach deepens the understanding of earthquake phenomena in Indonesia. The methodology employs advanced tools such as Microsoft Excel and ArcGIS Online Map for data processing, alongside Web AppBuilder for ArcGIS for visualization. This combination exemplifies a modern approach to geo-visual analytics workflow, ensuring the research remains relevant to contemporary applications in the field. Incorporating time-series animations in the dashboard allows users to observe fluctuations in earthquake activity over time. This feature is particularly useful for identifying trends and patterns in seismic events, aiding disaster preparedness and response. The results include a set of maps and graphics that serve as interactive visual interfaces. These tools are designed to enhance understanding of the relationships among various variables and to demonstrate visual analytics techniques through diverse example scenarios of spatio-temporal earthquake data [50]. [51] used Power BI to analyze earthquake data in Türkiye between 3.12.2022 and 30.12.2023. [52] proposed a study model showing how to improve earthquake prediction systems by combining IoT sensors with a Bidirectional Gated Recurrent Unit (Bi-GRU) machine learning model. Our study is based on digital data concerning significant earthquakes (those exceeding a magnitude of 5 on the Richter scale) and huge earthquakes (those exceeding a magnitude of 7 on the Richter scale). The data were obtained from USGS and include a comprehensive record of large-scale seismic events from 1843 to April 2025. The dataset can be accessed at the following link: earthquake.usgs.gov/earthquakes/search. PowerBI is used for data analytics and visualization to facilitate the analysis of global large-scale earthquake data, reflecting their evolution, impact, and recurring locations. Figure 1 presents a clear visual workflow illustrating the step-by-step process from data collection through prediction and visualization.

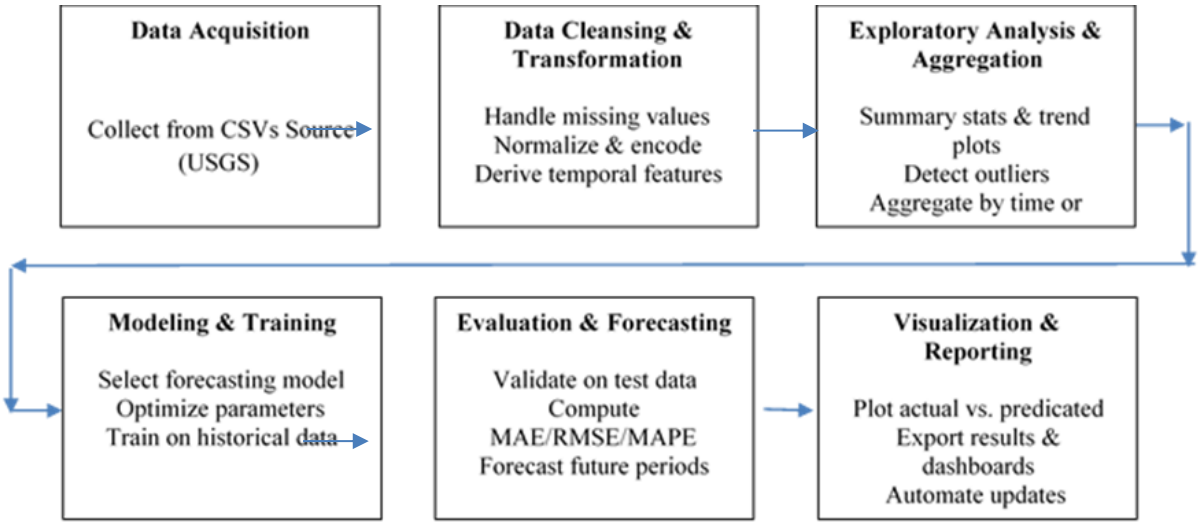


Figure 1. *The study workflow illustrating the step-by-step process from data collection through prediction aand visualization*

Evolution of the Number of Significant and Huge Earthquakes Annually

From 1800 to April 2025, more than 100,000 significant earthquakes occurred worldwide, affecting roughly 162 countries. About 2% of these, or around 2000 events, were classified as huge earthquakes (magnitude 7+). The analysis shows a decadal increase of over 45 significant earthquakes. Since 1973, there has been a marked rise in significant earthquakes. Instrumental records began in the late 18th century, so earlier events are estimates (Figure 2). The 1960 Great Chilean earthquake remains the largest on record. Figure 3 shows the number of significant earthquakes worldwide by year. A large increase of significant earthquakes was observed after 1973.

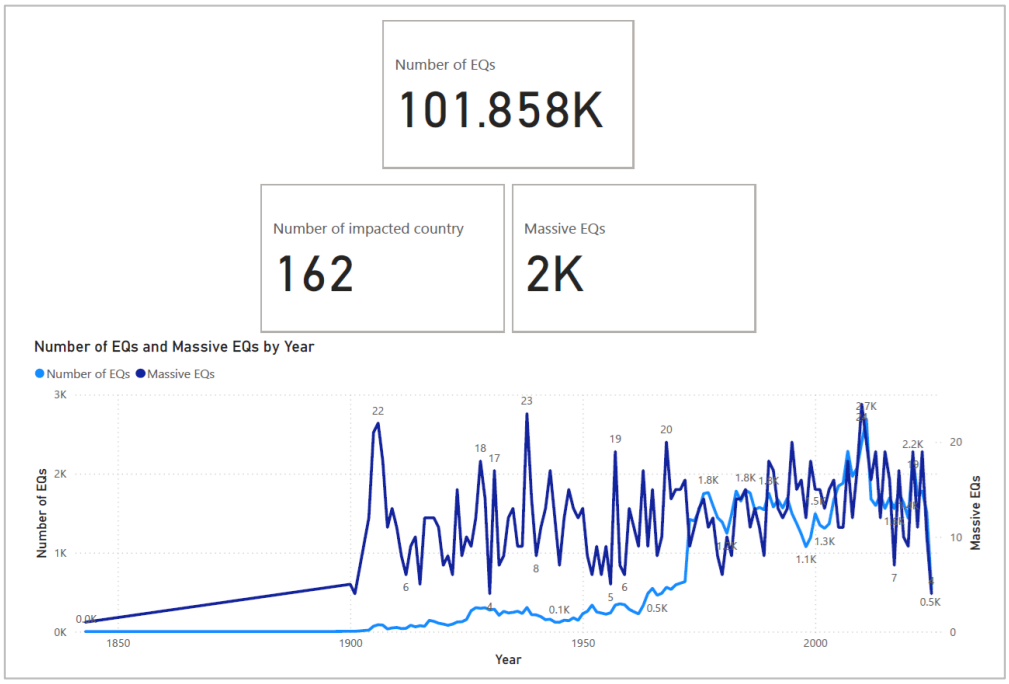


Figure 2. *Evolution of the annual number of significant and huge earthquakes annually; about 2000 earthquakes were classified as huge (magnitude 7+)*

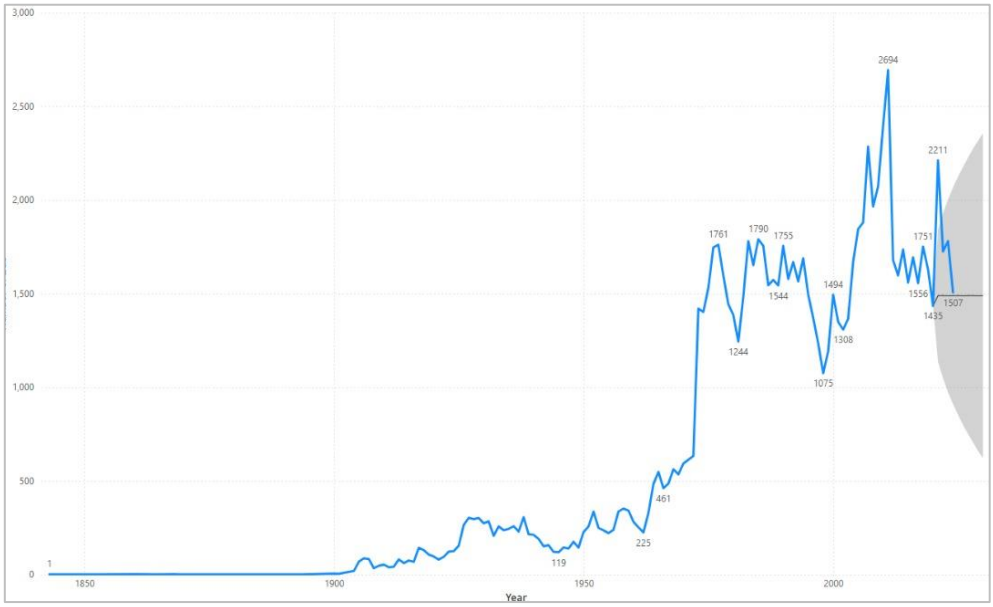


Figure 3. *Evolution of the annual number of the significant earthquakes (magnitude 5+) worldwide by year; a large increase of significant earthquakes was observed after 1973*

Table 3 summarises the statistical metrics values (MAE, RMSE, SMAPE) for time-series analytical method to forecast future occurrences of significant earthquakes worldwide.

Table 3. *Statistical metrics values*

Metric	Value
MAE	532.05
RMSE	630.47
SMAPE	35.44%

USGS estimates that since 1843 there have been an average of 18 huge earthquakes (magnitude 7.0–7.9) and one great earthquake (magnitude 8.0 or greater) per century, and that this average has remained relatively stable (Figure 4).

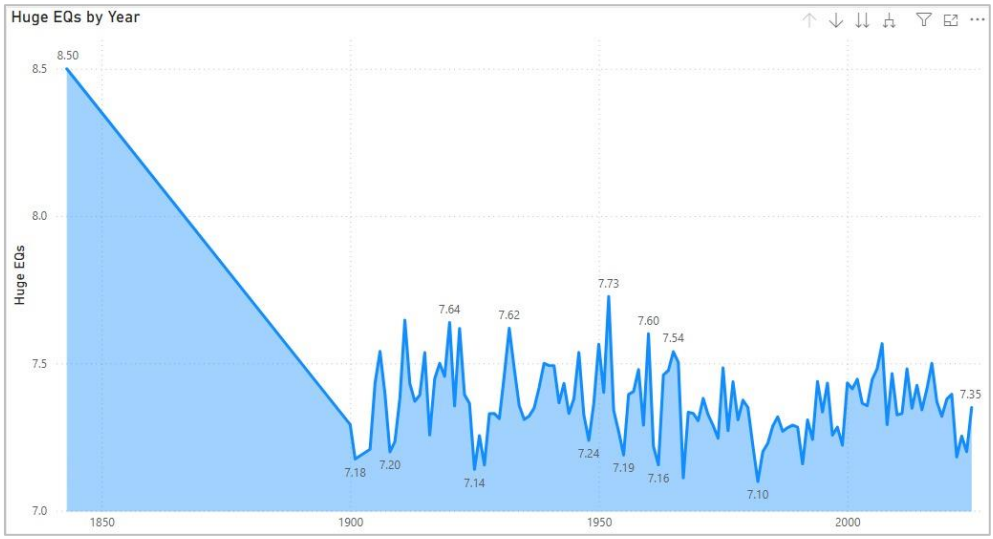


Figure 4. *Evolution of the number of huge (magnitude 7+) earthquakes worldwide per century; this average has remained relatively stable*

Largest Earthquakes Associated with Locations and Magnitudes

Figure 5 shows the occurrence of the ten most catastrophic earthquakes (with magnitudes exceeding 9 on the Richter scale) from 1843 to 2025. The 1960 Valdivia earthquake, registering a magnitude of 9.5, remains the most powerful earthquake ever documented. It struck off the coast of Chile, causing widespread devastation. The quake generated colossal tsunamis with waves reaching up to 25 meters, severely impacting the Chilean coastline and displacing around two million people. This event underscores the catastrophic potential of such seismic activity. Other regions frequently affected by similarly intense earthquakes include the United States, Japan, Ecuador, and Indonesia. These areas share their location along active tectonic plate boundaries, which makes them susceptible to frequent and severe earthquakes.

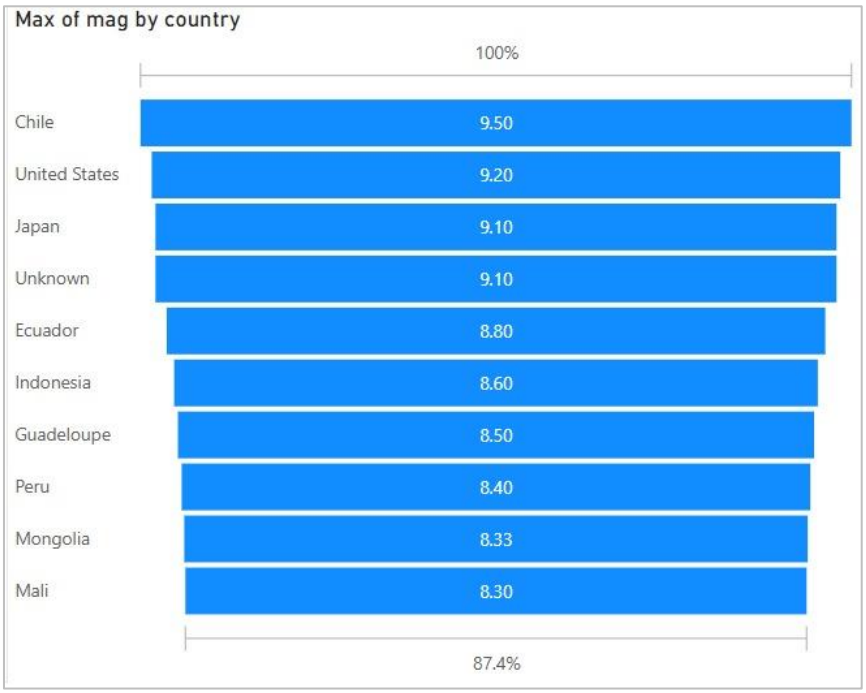


Figure 5*. The ten most catastrophic earthquakes associated with locations and magnitudes; the 1960 Valdivia earthquake, registering a magnitude of 9.5, remains the most powerful earthquake ever documented*

Number of Significant Earthquakes by Month

The provided data outlines the occurrence of significant earthquakes by month. Figure 6 indicates that March is the most active month, with 9330 significant earthquakes, followed by December with 8970 significant earthquakes, while June is the least active month with 7990 significant earthquakes.



Figure 6. *Number of significant earthquakes by month; March is the most active month, with 9330 significant earthquakes*

Average Magnitude (Mag) of Significant Earthquakes by Year

Figure 7 clearly illustrates a declining trend in average earthquake magnitudes since 1843, along with a possible correlation between decreasing magnitudes and rising earthquake frequency. The average magnitude of significant earthquakes decreased from 8.5 mag in 1843 to 5.84 in 1973 and further to 5.31 mag in 2023. This implies a shift towards more frequent but less intense seismic events. A notable decline in earthquake magnitude began in 1972, reaching 5.27 mag, even as their frequency increased from that point onward (Figure 2).



Figure 7. *Average magnitude (mag) of significant earthquakes by year; it illustrates a declining trend in average earthquake magnitudes since 1843, along with a possible correlation between decreasing magnitudes and rising earthquake frequency*

Average of Magnitude (mag) and Average of Depth of Significant Earthquakes by Year

Since 1843, an inverse relationship has been observed: as the average magnitude of significant earthquakes has decreased, their average depth has increased. The maximum depth was recorded in 1979. Although the provided snippets do not explicitly detail this relationship, they offer supporting context. Earthquake magnitude is related to the fault area; larger faults produce larger quakes. The intensity of shaking diminishes with distance from the source. The U.S. Geological Survey (USGS) data indicates that there have been approximately 18 huge earthquakes each century. This information suggests that while the frequency of major earthquakes has remained relatively stable, the average magnitude may have decreased, with depths varying over time (Figure 8).

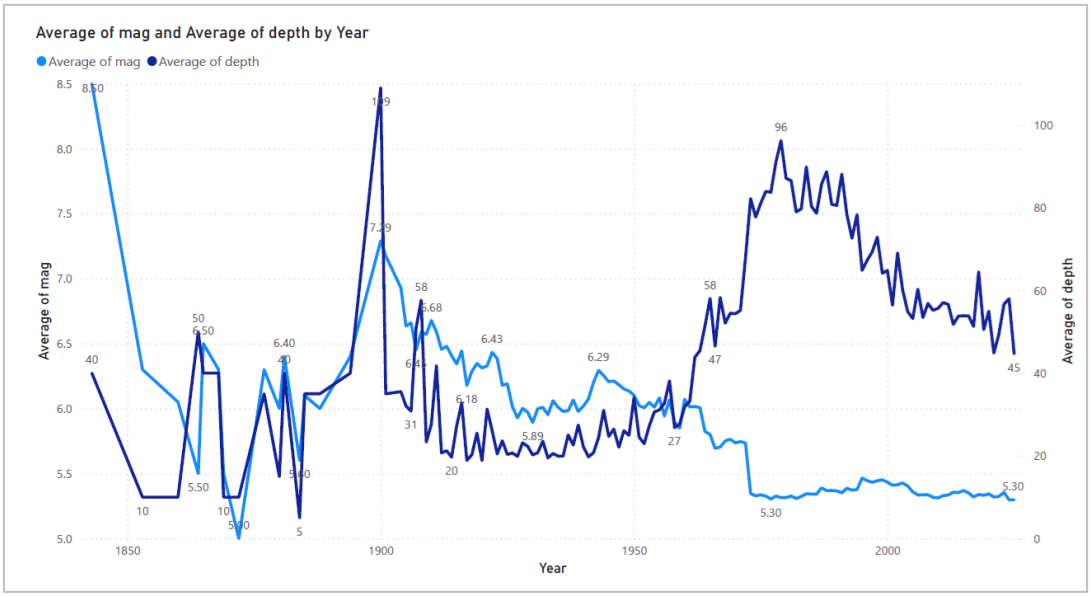


Figure 8. *Average of magnitude (mag) and average depth of significant earthquakes by year; it illustrates an inverse relationship: as the average magnitude of significant earthquakes has decreased, their average depth has increased*

Evolution of the Number of Significant Earthquakes in Japan

The implemented BI solution can classify and forecast data for any seismically active region globally. Figure 9 illustrates the trend in the number of significant earthquakes in Japan, a country that experiences daily seismic activity. The significant rise to 873 significant earthquakes in Japan in 2011 resulted from the devastating magnitude 9.0 earthquake on March 11, 2011. This event triggered a massive tsunami and more than 1,000 aftershocks, some exceeding magnitude 7.0. The high number of significant earthquakes in 2011 does not indicate a sustained trend but is instead a direct consequence of the energy released and the subsequent geological adjustments following the initial, powerful quake. The 2011 earthquake and its aftershocks greatly exceeded the annual average from 1900 to 2025 due to the immense energy released. Figure 8 also provides a forecast for the anticipated number of significant earthquakes in the coming years using the exponential smoothing algorithm. The notable increase to 873 significant earthquakes heavily influences the forecast.

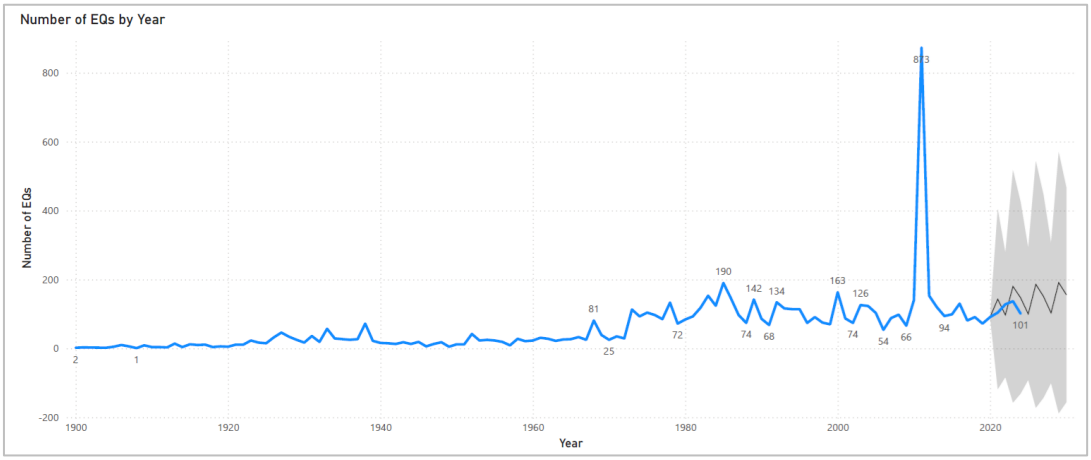


Figure 9. *Evolution of the number of significant earthquakes in Japan; the significant rise to 873 significant earthquakes in Japan in 2011 resulted from the devastating magnitude 9.0 earthquake on March 11, 2011*

Conclusion

In this study, we emphasize the importance of developing intelligent systems capable of analyzing vast amounts of digitally recorded earthquake data since 1843, using data mining and business intelligence tools. Such systems can significantly enhance decision-making and support the development of policies and procedures for earthquake prediction, including forecasting their magnitude, depth, and location. Our study highlights the value of creating a global data warehouse that consolidates earthquake records from accredited international sources. By integrating business intelligence tools, we aim to present findings through dashboards and visualizations.

Through this study, we have successfully achieved the following objectives:

- A comprehensive examination of the leading machine learning techniques, data mining tools, and business intelligence solutions in the field of seismology.
- Development of a specialized data warehouse for significant earthquakes, capable of analyzing both historical and real-time data.
- Generation of analytics for significant earthquake data using data mining tools and business intelligence solutions.

The findings of this study can be applied to correlate analyses of significant earthquake data across seismically active regions worldwide, with the aim of improving earthquake prediction accuracy using machine learning methods. This can be achieved by integrating diverse data sources such as satellite imagery and social media, and by developing intelligent algorithms that enhance the precision and reliability of earthquake forecasts. Ultimately, these advances could enable real-time earthquake monitoring systems capable of issuing early warnings based on observed indicators. Earthquake data analytics can provide emergency response agencies with timely, actionable intelligence to improve preparedness, enhance real-time response, and guide long-term recovery. By leveraging data from seismic sensors, satellite imagery, social media, and other sources, agencies can make data-driven

decisions to reduce causes and limit damage. Additionally, earthquake data analytics provides urban planners with vital information for creating more resilient, safer cities. By integrating data on seismic activity, ground conditions, and structural vulnerabilities, planners can shape land-use policies, reinforce infrastructure, and improve disaster response and recovery. Finally, such analytics give policymakers actionable insights for hazard mitigation, urban planning, emergency response, and financial risk management. By applying statistical and machine-learning models, governments can shift from reactive strategies to proactive, evidence-based policies that strengthen community resilience to seismic events.

Acknowledgements

Funding/Financial Disclosure The author has not received any financial support for the research, authorship, or publication of this study.

Ethics Committee Approval and Permissions The study does not require ethics committee permission or any special permission.

Conflict of Interests There is no conflict of interest.

Authors Contribution The authors read and approved the final manuscript

References

[1] Galkina, A., & Grafeeva, N. (2019). Machine learning methods for earthquake prediction: a survey. In: *Proceedings of the Fourth Conference on Software Engineering and Information Management*, pp. 25-32. Saint Petersburg, Russia.

[2] Gürsoy, G., Varol, A., & Nasab, A. (2023, May 11-12). *Importance of Machine Learning and Deep Learning Algorithms in Earthquake Prediction: A Review* [Conference presentation]. The 11th International Symposium on Digital Forensics and Security (ISDFS), Chattanooga, TN, USA. https://doi.org/10.1109/ISDFS58141.2023.10131766

[3] Xiong, P., Tong, L., Zhang, K., Shen, X., Battiston, R., Ouzounov, D., Iuppa, R., Crookes, D., Long, C., & Zhou, H. (2021). Towards advancing the earthquake forecasting by machine learning of satellite data. *Science of the Total Environment*, *771*, 145256. https://doi.org/10.1016/j.scitotenv.2021.145256

[4] Banna, M. H. A., Taher, K. A., Kaiser, M. S., Mahmud, M., Rahman, M. S., Sanwar Hosen, A. S. M., & Cho, G. H. (2020). Application of artificial ıntelligence in predicting earthquakes: state-of-the-art and future challenges. *IEEE Access*, *8,* 192880-192923. https://doi.org/10.1109/ACCESS.2020.3029859

[5] Nandwani, D. T., & Buradkar, V. (2022). Earthquake damage prediction using machine learning. *International Journal of Creative Research Thoughts, 10*(7), 206-211.

[6] Saleem, A. K., & Rashid, A. (2023). Applications of machine learning for earthquake prediction: A review. In *AIP Conference Proceedings: Al-Kadhum 2nd International Conference on Modern Applications of Information and Communication Technology (MAICT),* Baghdad, Iraq. 2591(1), 030042. https://doi.org/10.1063/5.0119623

[7] Alnoukari, M., Alhawasli, H. A., Abd Alnafea, H., & Zamreek, A. J. (2012). Business Intelligence:

Body of Knowledge. In: El Sheikh, A., Alnoukari, M. (eds.) Business Intelligence and Agile Methodologies for Knowledge-Based Organizations: Cross-Disciplinary Application, IGI Global, USA, pp. 1-13.

[8] Otari, G. V., & Kulkarni, Dr. R. V. (2012). A review of application of data mining in earthquake prediction. *International Journal of Computer Science and Information Technologies (IJCSIT), 3*(2), 3570-3574.

[9] McBrearty, I., & Beroza, G. (2023). Earthquake phase association with graph neural networks. *arXiv:2209.07086*. https://doi.org/10.1785/0120220182

[10] Zhang, Y., & Gao, S. S. (2024). Using convolutional neural network to determine time window for analyzing local shear-wave splitting measurements. *Seismological Research Letters, 95*(6), 3626–3632. https://doi.org/10.1785/0220230410

[11] Jarah, N. B., Alasadi, A. H. H., & Hashim, K. M. (2024). A new algorithm for earthquake prediction using machine learning methods. *Journal of Computer Science*, *20*(2), 150-156. https://doi.org/10.3844/jcssp.2024.150.156

[12] Chittora, P., Chakrabarti, T., Debnath. P., Gupta, A., Chakrabarti, P., Praveen, S. P., Margala, M., & Elngar, A. A. (2022). Experimental analysis of earthquake prediction using machine learning classifiers, curve fitting, and neural modeling. *Research Square*. https://doi.org/10.21203/rs.3.rs-1896823/v1

[13] Hoque, A., Raj, J., & Saha, Dr. A. (2018, December 14-15). *Approaches of Earthquake Magnitude Prediction using machine learning techniques* [Conference presentation]. International Conference on Computational Intelligence & IoT (ICCIIoT), Tripura, India. https://icciiot2018.iaasse.org/

[14] Bangar, M. H., Gupta, D., Gaikwad, S., Marekar, B., & Patil, J. (2020). Earthquake prediction using machine learning algorithm, *International Journal of Recent Technologies and Engineering*, *8*(6), 4684-4688. https://doi.org/10.35940/ijrte.E9110.018620

[15] Zaidoun, A., Diko, F., & Alnoukari, M. (2008). Enhancing education quality assurance using ınformation systems- QAAS system. *In IEEE International Symposium on Information Technology (ITSIM 08)*, Kuala Lumpur, Malaysia, pp. 1-6. https://doi.org/10.1109/ITSIM.2008.4631579.

[16] Alnoukari, M., Razouk, R., & Hanano, A. (2016). BSC-SI: A framework for ıntegrating strategic ıntelligence in corporate strategic management. *International Journal of Strategic Information Technology and Applications (IJSITA)*, *7*(1), 32-44. https://doi.org/10.4018/IJSITA.2016010103

[17] Lessy, D. F., Avorizano, A., & Hasan, F. N. (2022). Penerapan business ıntelligence untuk menganalisa data gempa bumi di ındonesia menggunakan tableau public. *Jurnal Sistem Komputer dan Informatika (JSON)*, *4*(2), 302. https://doi.org/10.30865/json.v4i2.5316

[18] Mer, S., Saxena, S., & Pundir, A. (2024). Unveiling Earthquake Dynamics: A Comprehensive Data Analytics and LSTM-Based Prediction Model for Enhanced Seismic Forecasting. In *IEEE 5th India Council International Subsections Conference (INDISCON),* Chandigarh, India, pp. 1-6, https://doi.org/10.1109/INDISCON62179.2024.10744325.

[19] Cremen, G., Seville, E., & Baker, J. (2020). Modeling post-earthquake business recovery time: an analytical framework. *International Journal of Disaster Risk Reduction, 42,* 101328. https://doi.org/10.1016/j.ijdrr.2019.101328

[20] Andrade-Arenas, L., & Yactayo-Arias, C. (2024). Seismic trend analysis: a data mining approach for pattern prediction. *IAES International Journal of AI*, *13*(3), 2623-2634.

http://doi.org/10.11591/ijai.v13.i3.pp2623-2634

[21] Zakeri, N. S. S., & Pashazadeh, S. (2015). Data Mining Techniques on Earthquake Data: Recent Data Mining Approaches. In M. Usman (Ed.), Improving Knowledge Discovery through the Integration of Data Mining Techniques (pp. 183-199). IGI Global Scientific Publishing. https://doi.org/10.4018/978-1-4666-8513-0.ch010

[22] Pasadana, I. A., Mawengkang, H., & Efendi, S. (2023). Predictive analysis of spatial data and time series to predict earthquake magnitudes by using data mining approach. *Randwick International of Social Science Journal*, *4*(1), 80–90. https://doi.org/10.47175/rissj.v4i1.607

[23] Diab-Montero, H. A., Stordal, A. S., van Leeuwen, P. J., & Vossepoel, F. C. (2025). Ensemble kalman, adaptive gaussian mixture, and particle flow filters for optimized earthquake forecasting. *Computers & Geosciences, 196,* 105836. https://doi.org/10.1016/j.cageo.2024.105836

[24] Özcan, M., & Peker, S. (2021). Designing a data warehouse for earthquake risk assessment of buildings: a case study for healthcare facilities. *Sakarya University Journal of Computer and Information Sciences*, *4*(1), 156-165. https://doi.org/10.35377/saucis.04.01.872729

[25] Matai, P., & Bhatia, A. (2024). Architecting for real - time analytics: Leveraging stream processing and data warehousing ıntegration. *International Journal of Science and Research (IJSR)*, *13*(9), 1586-1590. https://doi.org/10.21275/sr24925170923

[26] Mizutani, K., Mitarai, H., Miyazaki, K., Kumano, S., & Yamasaki, T (2024). Data-driven prediction of seismic ıntensity distributions featuring hybrid classification-regression models. *arXiv:2402.02150*. https://doi.org/10.48550/arXiv.2402.02150

[27] Bhujang, R. K., & Kotagi, V. (2023). Earth Quack Prediction of Southern California Using Unsupervised Clustering Based Ensemble Deep Learning Model. In *International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIE)*, Ballari, India, pp. 1-6, https://doi.org/10.1109/AIKIIE60097.2023.10390335.

[28] Vijaya Saraswathi, R., (2024). Machine learning-powered earthquake early warning system. *International Journal of Innovative Science and Research Technology*, *9*(6), 1492–1503. https://doi.org/10.38124/ijisrt/IJISRT24JUN1107

[29] Alabowsh, B., & Li, W. (2024, 14–19 Apr). An ınvestigation into the effectiveness of machine learning algorithms for earthquake magnitude prediction using seismic data. In *EGU General Assembly,* Vienna, Austria, EGU24-19288. https://doi.org/10.5194/egusphere-egu24-19288

[30] Xie, Y. (2024). Deep learning in earthquake engineering: A comprehensive review. *arXiv:2405.09021*, https://doi.org/10.48550/arXiv.2405.09021

[31] Akarsu, O. N., Akarsu, O., & Aydin, A. (2024). A bibliometric review of earthquake and machine learning research. *Civil Engineering Beyond Limits*, 1908, 1-10. https://doi.org/10.36937/cebel.2024.1908

[32] Arrowsmith, S. J., Trugman, D. T., Bergen, K. J., & Magnani, B. M. (2022, July 5). *The Big Data Revolution Unlocks New Opportunities for Seismology*. https://blog.smu.edu/dedmancollege/2022/07/05/the-big-data-revolution-unlocks-new-opportunities-for-seismology/

[33] Trugman, D. T., Fang, L., Ajo-Franklin, J. B., Nayak, A., & Li, Z. (2022). Preface to the focus section on big data problems in seismology. *Seismological Research Letters*, *93*(5), 2423–2425. https://doi.org/10.1785/0220220219

[34] Udegbe, E., Morgan, E., & Srinivasan, S. (2019). Big data analytics for seismic fracture ıdentification, using amplitude-based statistics. *Computer Geoscience, 23*, 1277–1291. https://doi.org/10.1007/s10596-019-09890-z

[35] Bhattacharjee, S., Rahim, L. B. A., Ramadhani, A. W., Midhunchakkravarthy & Midhunchakkravarthy, D. (2020). A study on seismic big data handling at seismic exploration industry. In: Peng, S. L., Son, L. H., Suseendran, G., Balaganesh, D. (eds) *Intelligent Computing and Innovation on Data Science*. Lecture Notes in Networks and Systems. vol. 118. pp. 421-429. https://doi.org/10.1007/978-981-15-3284-9_46

[36] Marketos, G., Theodoridis, Y., & Kalogeras, I. S. (2008). Seismological data warehousing and mining: A survey. *International Journal of Data Warehousing and Mining*, *4*(1), 1–16. https://doi.org/10.4018/jdwm.2008010101

Gerasimos, M., Yannis, T., & Ioannis, S. K. (2010). Seismological Data Warehousing and Mining: A Survey. In D. Taniar & L. Rusu (Eds.), Strategic Advancements in Utilizing Data Mining and Warehousing Technologies: New Concepts and Developments (pp. 22-37). IGI Global Scientific Publishing. https://doi.org/10.4018/978-1-60566-717-1.ch002

[37] Cornely, P. R., & Wang, J. (2023). Advancing earthquake prediction: A comprehensive review of data science techniques. In *6th International Conference on Computing and Big Data (ICCBD)*, Shanghai, China, pp. 9–16. https://doi.org/10.1109/ICCBD59843.2023.10607190

[38] Kollam, M., & Joshi, A. (2023). Earthquake forecasting using optimized levenberg–marquardt back-propagation neural network. *WSEAS Transactions on Computers*, *22,* 90–97. https://doi.org/10.37394/23205.2023.22.11

[39] Ye, Q., Xia, B., & Ren, Y. (2024). Towards a multimedia big data-driven approach for earthquake monitoring and forecasting early warning system. *Informatica*, *48*(9), 53-64. https://doi.org/10.31449/inf.v48i9.5798

[40] Chandana, K. S., Sandeep, U. S., Asritha, P., Mothukuri, R., & Reddy, M. J. (2024). Seismic magnitude forecasting through machine learning paradigms: A confluence of predictive models. *International Journal of Innovative Science and Research Technology*, *9*(6), 2606–2613. https://doi.org/10.38124/ijisrt/IJISRT24JUN2025

[41] Anbazhagu, U. V., Sonia, R., Grover, R. K., Banu, E. A., Jothikumar, C., & Sudhakar, M. (2024). AI and Machine Learning in Earthquake Prediction. In F. Smarandache & P. Majumder (Eds.), *Advances in Computer and Electrical Engineering Book Series*, pp. 1–32. IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-6875-6.ch001

[42] Gitis, V., & Derendyaev, A. (2019). Machine learning methods for seismic hazards forecast. *Geosciences, 9*(7), 308. https://doi.org/10.3390/geosciences9070308

[43] Arunadevi, B., Hussain, M, M., M. I., Lakshmi, R. M. M., R., & Das, K. S. (2022). Risk Prediction of Earthquakes using Machine Learning. In *3rd International Conference Electronic Systems, Signal Processing and Computing Technologies*, pp. 1589–1593. https://doi.org/10.1109/ICESC54411.2022.9885674.

[44] Surisetty, H. V., Gottimukkala, S. V., & Selvam, S. (2023). A Streaming Application for Real-Time Analytics of Seismic Data. *4th IEEE Global Conference for Advancement in Technology (GCAT)*, Bangalore, India. pp. 1-8. https://doi.org/10.1109/GCAT59970.2023.10353520

[45] Atsa'am, D. D., Gbaden, T., & Wario, R. (2023). A machine learning approach to formation of

earthquake categories using hierarchies of magnitude and consequence to guide emergency management. *Data Science and Management*, *6*(4), 208-213. https://doi.org/10.1016/j.dsm.2023.06.005

[46] Franke, M., Martin, H., Koch, S., & Kurzhals, K (2021). Visual analysis of spatio-temporal phenomena with 1D projections. *Computer Graphics Forum*, *40*(3), 335-347. https://doi.org/10.1111/cgf.14311

[47] Kulkarni, S., Phadke, M., Sawant, A., Patel, N., & Patil, O. (2025). Advancements in seismic data collection and analysis through machine learning. *Indonesian Journal of Electrical Engineering and Computer Science, 37*(3), 2058-2068. https://doi.org/10.11591/ijeecs.v37.i3.

[48] Zhu, W., Alvin, H., Robert, Y., Avoy, D., Mostafa, M. S., William, E., & Gregory, B. (2022). QuakeFlow: A scalable machine-learning-based earthquake monitoring workflow with cloud computing. *Geophysical Journal International, 232*(1), 684–693. https://doi.org/10.1093/gji/ggac355

[49] Mousavi, S. M., Sheng, Y., Weiqiang, Z., & Beroza, G. (2019). STanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI. In *IEEE Access*, vol. 7, pp. 179464-179476, https://doi.org/10.1109/ACCESS.2019.2947848.

[50] Susanta, F. F., Pratama, C., Aditya, T., Khomaini, A. F., & Abdillah, H. W. K. (2019). Geovisual analytics of spatio-temporal earthquake data in Indonesia. *Journal of Geospatial Information Science and Engineering, 2*(2), 185-194. https://doi.org/10.22146/jgise.51131

[51] Ernez, D., Akinci, Y., Batmaz, M., Alinca Ayyildiz, Y. (2025). Analysis and visualization of earthquake data with POWER BI in Türkiye: an evaluation from the perspective of education and guidance. *International Journal of Social Science*, *Innovation and Educational Technologies*, *6*(22), 45-63. http://dx.doi.org/10.54603/iss.225

[52] Rasool, K., Hussain., M., Rasool., U., Umar, G., Rusool, J., Maryam, M. (2025). Seismic data analysis and earthquake prediction with IoT sensors and SmartGRU model. *International Journal of Innovations in Science & Technology (IJIST), 7*(3)*,* 1376-1395. https://doi.org/10.33411/ijist/20257313761395