Integrating Data Mining and Seismic Modelling for Natural Hazard Assessment in Tectonic Zones

Mrudula Manish Gudadhe [1], Dr. Narendra Chaudhari [2]

[1,2] Department of Computer Science & Engineering, Mansarovar Global University, Sehore, MP, India.

ABSTRACT

In order to improve the evaluation of natural hazards in tectonically active zones, this research offers an integrated strategy that combines data mining approaches with seismic modeling. Conventional seismic models don't always do a good job of representing nonlinear connections and dynamic tectonic activity, especially with the growing amount and complexity of complex geographic data. To improve forecast accuracy and early warning capabilities, data mining may be used as a nonparametric analytical strategy to uncover hidden patterns and correlations within vast, diverse datasets. Data mining helps find seismic precursors, estimate event magnitudes, and identify high-risk zones by using machine learning methods including clustering, regression, and classification. This hybrid approach provides a more thorough comprehension of possible hazards when combined with probabilistic and deterministic seismic models. Dynamic risk mapping, multi-hazard interaction modeling, and real-time hazard monitoring are the main points of the research. It further shows how these methods, when combined, can build stronger infrastructure and make better policy choices. The integrated model has great promise for enhancing disaster preparation and risk mitigation measures in earthquake-prone locations, despite obstacles such as data quality, computing needs, and multidisciplinary collaboration. Findings from this study point to data-driven, adaptable approaches as the way forward for seismic risk assessment.

Keywords: *Data Mining, Seismic Modeling, Natural Hazard Assessment, Tectonic Zones, Earthquake Prediction.*

I. INTRODUCTION

In tectonically active zones, natural disasters, especially earthquakes, continue to be a major concern for people, buildings, and ecosystems. For successful risk assessment and mitigation, rigorous, predictive, and multi-disciplinary techniques are required for earthquakes and its related hazards, which are complex and nonlinear processes. When it comes to understanding tectonic activity and hazard predictions, traditional seismic modeling techniques—which are mostly based on geophysics, geology, and seismology—have been invaluable. Nevertheless, a chance to improve these traditional approaches by incorporating data mining techniques has arisen due to the growing availability of large-scale environmental, geographical, and seismic data. Data mining is a branch of AI and computer science concerned with discovering useful correlations and patterns in massive databases. Information collected from satellite imagery, sensor networks, historical seismic recordings, GPS data, and remote sensing technologies is particularly significant for natural hazard assessment in tectonic zones due to its capacity to manage large amounts of diverse data.

Data mining helps forecast seismic event magnitudes and frequencies by identifying antecedents to seismic events, delineating high-risk zones, and employing methods including clustering, regression analysis, classification, and machine learning algorithms. Because of their potential inability to represent the multi-scalar, dynamic character of tectonic processes, conventional seismic models may have their inadequacies addressed by this integration. A new way of looking at natural disaster risk has emerged with the combination of data mining and seismic modeling.

Both probabilistic and deterministic methods of seismic modeling are based on mathematical and physical models of fault mechanics, ground motion propagation, and the interaction of tectonic plates. Although these models play an essential role in comprehending seismicity's physics, their precision is often limited by inputs that are either missing data or out of date. When used in conjunction with seismic models, data mining may improve their prediction capacity and flexibility by adding real-time and historical statistics. In order to create dynamic seismic models, machine learning algorithms may sift through mountains of seismic data in search of irregularities and subtle patterns that might indicate an impending big earthquake. The ability for early warning systems and real-time danger monitoring is one of the main benefits of this integration.

Algorithms for data mining may continually examine data coming in from seismic stations, satellite feeds, and ground sensors; this allows for speedier community and government reactions by alerting on abnormal behavior. This fast analysis and forecast system might greatly reduce the effects of disasters in tectonic zones, especially those located in developing or heavily inhabited areas.

Urban planners and legislators may build resilient infrastructure and communities with the use of dynamic hazard maps that represent real-time tectonic pressures and historical event patterns, made possible by data mining. In tectonic zones, where risks are frequent and amplified by other environmental weaknesses, this multidisciplinary strategy is of the utmost importance. Seismicity, floods, landslides, and other climate-related hazards are all too common in areas located within the Himalayan Belt or the Pacific Ring of Fire. To get a more complete picture, showing how several dangers are dependent on one another, it is best to combine seismic models with multi-hazard data mining techniques.

The development of integrated disaster risk reduction (DRR) plans is facilitated by this comprehensive view. Data mining and seismic modeling integration has great promise, but it is not without its share of obstacles. Important obstacles to successful implementation include inadequate data or lack of availability, computational constraints, uninterpretable models, and the need for multidisciplinary knowledge. Additionally, there is an urgent need for global partnerships to develop open-access platforms that support this kind of integrative research, standardize procedures, and exchange data.

It is important to consider the ethical implications of data privacy, particularly when working with data collected from crowds or from mobile devices. One exciting new area for natural hazard assessment is the combination of data mining with seismic modeling. To lessen the impact of seismic hazards, it provides real-time monitoring, improved prediction capabilities, and dynamic risk mapping. Innovative and integrative techniques like these are crucial for protecting populations and fostering sustainable development in areas prone to hazards, since tectonic activity is a constant and unpredictable natural force.

Data Mining

Data mining is a powerful analytical process designed to explore large amounts of data in search of consistent patterns, trends, anomalies, and relationships that can be used to make valid predictions or informed decisions. As the volume of digital data grows exponentially across industries and domains, data mining has become an essential tool for extracting valuable insights from raw data. The term "data mining" is somewhat misleading, as it suggests the extraction of data, whereas the process actually involves analyzing data to discover patterns and knowledge. Data mining is one of the steps in the broader process of Knowledge Discovery in Databases (KDD), which includes data selection, preprocessing, transformation, data mining itself, and interpretation or evaluation of the discovered patterns. Data mining integrates principles from multiple fields including statistics, machine learning, artificial intelligence, and database systems to uncover useful information from vast datasets. Its goal is not only to find patterns in data but to ensure these patterns are useful and can support decision-making in real-world contexts. Various techniques are employed in data mining such as classification, clustering, regression, association rule learning, anomaly detection, and sequential pattern mining. Classification involves predicting a categorical label, such as whether an email is spam or not. Clustering, on the other hand, groups data points that are similar to each other without prior labels, useful in market segmentation and customer profiling. Regression is used to predict continuous values like house prices or stock prices. Association rule mining identifies relationships between variables in large databases, commonly used in market basket analysis to find products frequently bought together. Anomaly detection identifies unusual data records, often used in fraud detection and network security. These techniques help organizations to make proactive, knowledge-driven decisions and improve operational efficiency.

In the business world, data mining is widely used for customer relationship management, risk assessment, fraud detection, and targeted marketing. For instance, by analyzing customer purchase patterns, companies can tailor their advertising strategies, improve customer service, and increase sales. Banks and financial institutions use data mining to evaluate credit risks, detect suspicious transactions, and improve investment strategies. In the healthcare sector, data mining is used for diagnosing diseases, identifying treatment effectiveness, and managing hospital resources more effectively. Medical data mining helps in identifying risk factors for diseases and predicting outcomes based on patient history. Educational institutions use data mining to analyze student performance, identify at-risk students, and personalize learning experiences. Government agencies utilize data mining for crime detection, tax fraud detection, and public health monitoring. In e-commerce, data mining powers recommendation systems by analyzing customer behavior and preferences, thus enhancing user experience and sales. Social media platforms use data mining to understand user interests, detect fake accounts, and deliver personalized content and advertisements. In the field of scientific research, data mining helps in analyzing large datasets generated from experiments and simulations, aiding in discovery and innovation. Despite its many advantages, data mining is not without challenges. One of the primary challenges is data quality. The results of data mining are only as good as the data fed into the system. Incomplete, noisy, or inconsistent data can lead to inaccurate conclusions. Data privacy and security is another major concern, especially when dealing with sensitive information such as personal, financial, or medical records. The misuse of data mining techniques can lead to ethical issues, such as profiling or discrimination.

II. LITERATURE REVIEW

Priyanto, Dadang et al., (2022) so yet, scientists have been unable to pin down earthquakes' exact causes or predict when they will occur. A number of approaches have been devised, one of which pertains to data mining. These include clustering, fuzzy modeling, support vector regression, hybrid neural networks, and a host of others. A suitable strategy is required to produce best findings in earthquake research, which contains unclear parameters. Typically, parametric and non-parametric approaches are used to classify various predictive data mining techniques. The non-parametric approach used in this work is based on the MARS algorithm's backward step, which is conic multivariate adaptive regression spline (CMARS). Parameter testing and analysis yielded a mathematical model with 16 basis functions (BF) in this work; 12 of these basis functions were considered important in building the model, while 4 were deemed unimportant. The magnitude is 31.1%, the site temperature is 5.5%, the depth is 3.5%, and the epicenter distance is 100% according to the degree of variable contribution. Prediction analyses have shown that Malaka, Genggelang, Pemenang, Tanjung, Tegal Maja, Senggigi, and Mangsit are the places of Lombok most at risk from earthquakes.

Priyanto, Dadang et al., (2020) Data mining is the practice of discovering insights and patterns in large datasets. There are two main schools of thought when it comes to data mining: descriptive and predictive. The Classification and Regression function is one of several Math functions that may be used in data mining. An often-used statistical tool for studying and modeling correlations between variables, Regression Analysis (also termed Prediction Analysis) is ubiquitous in the field. The estimation of the regression curve may be accomplished by the study of Nonparametric Regression in regression. Among the many popular non-parametric regression methods, MARS stands out. When the Linear Regression approach has flaws, the MARS technique may fix them. A novel approach named CMARS (Conic Multivariate Adaptive Regression Splines) was developed by combining the CQP quadratic programming framework (CQP) from MARS with a stepwise backward algorithm. Modeling high-dimensional data with nonlinear structures is within the scope of the CMARS approach. Analyzing earthquake forecasts, particularly in Lombok, West Nusa Tenggara, might benefit from the CMARS model's adaptability. The dependent variable, Peak Ground Acceleration (PGA), yields statistically significant findings when a mathematical model including four independent factors is established. A total of 100%, 31.1% by magnitude, 5.5% by incident site temperature, and 3.5% by depth make up the independent variables.

Zhang, Wengang et al., (2015) it is not feasible to depend solely on simulation for the goal of design optimization due to the enormous computational costs of running complex numerical analyses like finite element simulations. This is because many geotechnical problems are extremely nonlinear and multivariate, even though the processing speed and memory of affordable computers are increasing rapidly. In order to keep costs down, surrogate models, which are also called meta-models, are built and used instead of the real numerical simulation models. For a more trustworthy surrogate model, it's best to use design variables with large ranges. For this reason, it is preferable to use meta-modeling approaches that can handle multivariate issues. In order to approximate the connection between inputs and outputs with huge data, this work investigates the usage of multivariate adaptive regression splines (MARS), a very straightforward nonparametric regression approach. An extensive explanation of the MARS approach and its related operations is provided first. To showcase MARS's function approximation skills and its effectiveness in handling multivariate issues with enormous data sets, two complex geotechnical problems are next given. The MARS method can estimate the contributions of input variables and provide simple, accurate, and easily interpretable models, as shown in this study.

Zhang, Wengang & Goh, Anthony. (2014). The loads from the superstructure are transferred onto the more rigid soils or rocks by means of piles, which are long and thin structural components. The piling hammer's impact causes compression and tension strains in driven piles. Therefore, it is crucial to ensure that the pile's strength is enough to withstand the pressures generated by the pile hammer's impact as a design factor. There is no exact analytical answer to pile drivability with respect to the phenomena involved because of its complexity. The nonlinear interactions and linkages between the system's predictors and dependent responses may be best mapped by neural networks when numerical hypothetical outcomes or measurable data are available. A further advantage over other computational tools is that it is not necessary to assume any mathematical link between the independent and dependent variables. However, others have pointed out that neural networks take a long time to train due to the fact that the best configuration isn't known in advance. Rather than using neural networks, this study looks at using multivariate adaptive regression splines (MARS), a relatively simple nonparametric regression algorithm, to approximatively estimate the input-dependent response relationship and to mathematically interpret the relationship between the different parameters. The drivability of piles is evaluated in this study using the MARS and Back propagation neural network (BPNN) models for predicting MTS, BPF, and Maximum compressive stresses (MCS). The model construction and comparison of BPNN and MARS predictions are conducted using a database including over 4,000 piles.

Yerlikaya-Özkurt, Fatma et al., (2014) For a given distance from the epicenter of an earthquake, the peak ground reaction may be estimated using empirical connections known as Ground Motion Prediction Equations (GMPEs). They establish a relationship between the amplitude and depth of an earthquake, the circumstances at the recording location, the kind of earthquake's source, and the ground's peak reactions. This paper presents a novel GMPE that is derived from an existing dataset using a prediction technique called Conic Multivariate Adaptive Regression Splines (CMARS). Conic quadratic programming is the unique continuous optimization method upon which CMARS is built. The employment of interior point techniques is made possible by the fact that these convex optimization problems are highly organized, looking like linear algorithms. We run the CMARS algorithm using Turkey's robust ground motion datasets. Three different GMPEs are used to compare the results. When it comes to predicting ground motion, CMARS works well.

Güllü, Hamza. (2012). One of the main aspects that significantly impacts earthquake-induced structural damage is peak ground acceleration (PGA). Seismic hazard evaluations now often include PGA prediction and suitable ground motion model selection as key topics. In this research, we provide an application that uses genetic expression programming (GEP), a novel prediction tool, and the traditional regression approach to forecast the PGA. Then, we try to figure out which ground motion models based on GEP and regression are the best. In order to rank the ground-motion models for seismic hazard analysis in regions of moderate seismicity, specifically in the case of rock motion, the prediction performances were compared. In order to derive the candidate ground motion models of PGA attenuation equations, the appropriately structured data from the Turkish earthquake was used. Key attenuation properties, Turkish attenuation equations, and case records of strong ground motion data in Turkey were used to validate the LH technique, GEP, and regression models. Model validations are often successful for most PGA candidate models (GEP and regressions) listed as strong qualifiers (class A and B) according to the LH approach, while models ranked lower (class C) typically fail.

Samui, Pijush & Kurup, Pradeep. (2012). In this paper, we look at the possibility of using MARS and LSSVM, which stand for multivariate adaptive regression spline, to forecast the OCR of clay deposits using data from Piezocone Penetration Tests (PCPT). The non-linear connections between the input and output variables are described by MARS using piece-wise linear segments. Using the regression approach, LSSVM is grounded on the idea of statistical learning. Modified cone resistance (q t), total stress vertical (σv), hydrostatic pore pressure (u 0), pore pressure at the cone tip (u 1), and the pore pressure immediately above the cone base (u 2) are the parameters that the models use as input. An error bar for the expected OCR is provided by the constructed LSSVM model. Predictions of OCR have also been made using equations. Traditional OCR prediction techniques have been compared to the performance of MARS and LSSVM models. The findings show that the MARS and LSSVM models that were suggested are good ones for determining OCR.

III. RESEARCH METHOD

Multivariate Adaptive Regression Spline (MARS)

To find the pattern of the association between the response variable and the predictor variable whose regression curve is unknown, the MARS technique is used. This nonparametric regression approach helps to address the difficulty of high-dimensional data. Parametric regression and nonparametric regression are the two main methods for completing predictions in data mining management.

As statistical tools, these two tacks see heavy application in studying and modeling inter-variable correlations. In order to address the problems with Recursive Partitioning Regression (RPR), the MARS approach can detect additive linear functions and generate continuous models at knots. In order to solve the MARS technique, the algorithm consists of two stages: the Forward Stepwise model and the Backward Stepwise model. Combining fundamental functions (BF), maximum interaction (MI), and minimal observation (MO) in the first stage means using the Forward Stepwise Algorithm. The study's predictor variables are depth, magnitude (Mw), and epicenter distance (Repi), whereas the response variable is Peak Ground Acceleration (PGA).

The second stage of the Backward Stepwise model is used for the purpose of simplifying the basis function (BF) that was produced from the Forward Stepwise stage. In the backward stepwise model stage, the basis function (BF) that contributes little or nothing to the response variable will be removed. The number of least squares of the remaining value will decrease as a result of this elimination operation.

Peak Ground Acceleration (PGA)

The highest acceleration of ground vibrations in a region induced by an earthquake is known as Peak Ground Acceleration (PGA). If the PGA value is high in one location, the earthquake's epicenter will likely take a major beating. Gravitational acceleration, abbreviated "gal," is the standard unit of measurement for PGAs.

Using the empirical computation of the Attenuation function is one technique to determine the PGA value. An area's susceptibility to ground vibrations, their amplitude, and the distance from the earthquake's epicenter are all determined by the attenuation function. The attenuation function is impacted by a number of variables, including the earthquake mechanism, the distance to the epicenter, and the state of the ground at the site.

Also, the study used prediction analysis, which included selecting and separating relevant factors for the Responsive and Predictor variables in advance. Within this research, the response variable 'PGA' is used, with depth, magnitude (Mw), and epicenter distance (Repi) serving as the predictor factors.

Data Collection

Data collected from earthquake catalogs is used in this investigation. The data was filtered with a magnitude greater than 4 Mw and accessed in October 2024. The reason for this is because an earthquake with a magnitude below 4 Mw is unlikely to produce any noticeable damage or even be felt at all. Over the course of 20 years, 105 recordings were compiled from earthquake catalogs, spanning from 4 to 5.5 Mw.

A selection mechanism is used to process the data, which requires an earthquake magnitude more than 4 Mw, a depth below 250 Km, and a center distance below 300 Km. Because it does not pose a threat, data that is not part of the provisions or ring will either be removed or not used. Predictive analysis makes use of the processed data. The magnitude, distance to the epicenter, and depth of the earthquake's core are the three factors that have been identified.

Eq. 6 was used in the Multivariate Adaptive Regression Spline (MARS) approach for earthquake prediction analysis, while Eq. 7 was utilized to find the lowest value of Generalized Cross Validation (GCV). Earthquake prediction using SPM 8 software included examining the parameter factor of the predictor-response relationship. A pair of algorithms, known as Forward Stepwise and Backward Stepwise, is used by MARS. Combining the maximum basis function with the maximum interaction and minimal observation (MO) is what the Forward Stepwise method finds.

The greatest possible basis function for a cross-multiplication of two linked or correlated variables. The Maximum Interaction (MI) defines the longest possible path through the basis function (BF) that may pass through the knot site, while the Minimum Observation (MO) determines the smallest possible value for the smoothing parameter, or the smallest possible distance between knots. And to make the constructed mathematical model functions even simpler, the Backward Stepwise technique is used.

The approach employs the Tikhonov Regularization technique, which imposes a penalty on mathematical models with too complicated functions, to reduce generalization mistakes. One way to tell whether a place is vulnerable to earthquakes is to look at its Peak Ground Acceleration (PGA). When an earthquake strikes a region with a high PGA value, it will cause significant damage. This research employs empirical calculations using the Joyner and Boore Attenuation functions to get the PGA value, which may be acquired either via accelerograph recordings or from other sources.

IV. RESULTS AND DISCUSSION

Results

Epicenter distance and Maximum Ground Acceleration (PGA) values were determined during data preprocessing, which formed the basis of the study's findings. The PGA value was determined using the Joyner and Boore attenuation functions. The MARS approach may be used to continue the computation and prediction analysis after the PGA value is known. Choosing the best model by testing the data and finding the least GCV value is the next step in getting the best MARS model.

The highest acceleration of ground vibrations in a region induced by an earthquake is known as Peak Ground Acceleration (PGA). If the PGA value is high in one location, the earthquake's epicenter will likely take a major beating. Gravitational Acceleration joules (PGA) are the standard units of measurement. Using the empirical computation of the Attenuation function is one technique to determine the PGA value. An area's susceptibility to ground vibrations, their amplitude, and the distance from the earthquake's epicenter are all determined by the attenuation function.

The attenuation function is impacted by a number of variables, including the earthquake mechanism, the distance to the epicenter, and the state of the ground at the site.

Also, the study used prediction analysis, which included selecting and separating relevant factors for the Responsive and Predictor variables in advance. Within this research, the response variable 'PGA' is used, with depth, magnitude (Mw), and epicenter distance (Repi) serving as the predictor factors. Finding out what happens when you use the right variables in your prediction analysis is possible. The training data for the MARS method's prediction analysis is derived from the outcomes of the Forward Stepwise and Backward Stepwise algorithms, which are based on a mix of BF, MI, and MO. Table following displays the results of the MARS regression using the training data.

Table 1: Results of Training Data

Parameter	Estimate	SE.	T-Value	P-Value
Constant	0.04286	0.00029	148.3594	0
Basis Function 1	-0.00131	0.00007	-18.98361	0
Basis Function 2	0.00151	0.00003	58.67444	0
Basis Function 3	0.02211	0.0011	20.06196	0
Basis Function 4	-0.02334	0.00054	-43.43563	0
Basis Function 5	0.00107	0.00008	13.60181	0

Basis Function 7	0.00038	0.00003	10.96367	0
Basis Function 9	0.00031	0.00004	6.99231	0
Basis Function 11	0.00039	0.00006	6.58793	0
F-STATISTIC = 5977.78679			S.E. OF REGRESSION = 0.00035	
P-VALUE = 0.00000			RESIDUAL SUM OF SQUARES = 0.00000	
[MDF, NDF] = [8, 23]			REGRESSION SUM OF SQUARES = 0.00577	

Testing and Analysis

A statistical analysis test is essential in predictive analysis for obtaining the results of hypothesis testing and determining the degree of significance. One way to determine the importance of a parameter is to use the significance level. In order to apply statistical analysis to find out how important parameters are in relation to the mathematical model's applicability, hypothesis testing is necessary. This study employs a partial regression coefficient test to evaluate mathematical model analysis.

Discussion

Table 1 shows that the response variable is influenced by parameters created using 11 basis functions: 1, 2, 3, 4, 5, 7, 9, and 11. Afterwards, the basic function is removed or omitted as it does not contribute to the answer variable. This includes basis functions 6, 8, and 10. A mathematical model, as shown in Formula below, may be used to determine the results of evaluating the data at the Backward Stepwise stage by simplifying the function.

$$Y = 0.042863 - 0.00130501 * BF1 + 0.00151234 * BF2 + 0.0221103 * BF3 - 0.0233377 * BF4 + 0.00106639 * BF5 + 0.000377886 * BF7 + 0.000305277 * BF9 + 0.000391561 * BF11;$$

MODEL PGA = BF1 BF2 BF3 BF4 BF5 BF7 BF9 BF11;

The following is the formula for Y ($_{P\ GA}$), which is the outcome of PGA Prediction analysis using the MARS model, with the contribution of each basis function (BF):

BF1 = max (0, REPI - 50.3651);
BF2 = max (0, 50.3651 - REPI);
BF3 = max (0, MW - 5.1);
BF4 = max (0, 5.1 - MW);
BF5 = max (0, MW - 4.8) * BF2;
BF7 = max (0, REPI - 41.4528) * BF4;
BF9 = max (0, REPI - 64.7445);
BF11 = max (0, REPI - 44.9069);

Table 2 shows the interplay of the predictor variable's contribution to the response variable, and it is based on the best MARS model. The predictor variables that impact PGA are determined by the MARS model in a sequential fashion according to the percentage of their contribution, which is determined by the smallest GCV value. These variables are the distance of the epicenter (Repi), the magnitude (Mw), and the depth (Depth).

Table 2: The Interactivity of Predictor Variable Contributions

Variable	Importance	-gcv
REPI	100.00000	0.00023
MW	73.80473	0.00012
DEPTH	0.00000	0.00000

Table 2 shows that the two most important factors in the PGA value—the depth (Depth)—contribute zero percent, while the epicenter distance (Repi) and magnitude (Mw) account for seventy-three percent and seventy-two percent, respectively. Figure 1 displays the test results for the three-dimensional graphs of the predictor variable's contribution to the response variable, which help to explain the description of the variable contributions of each predictor.

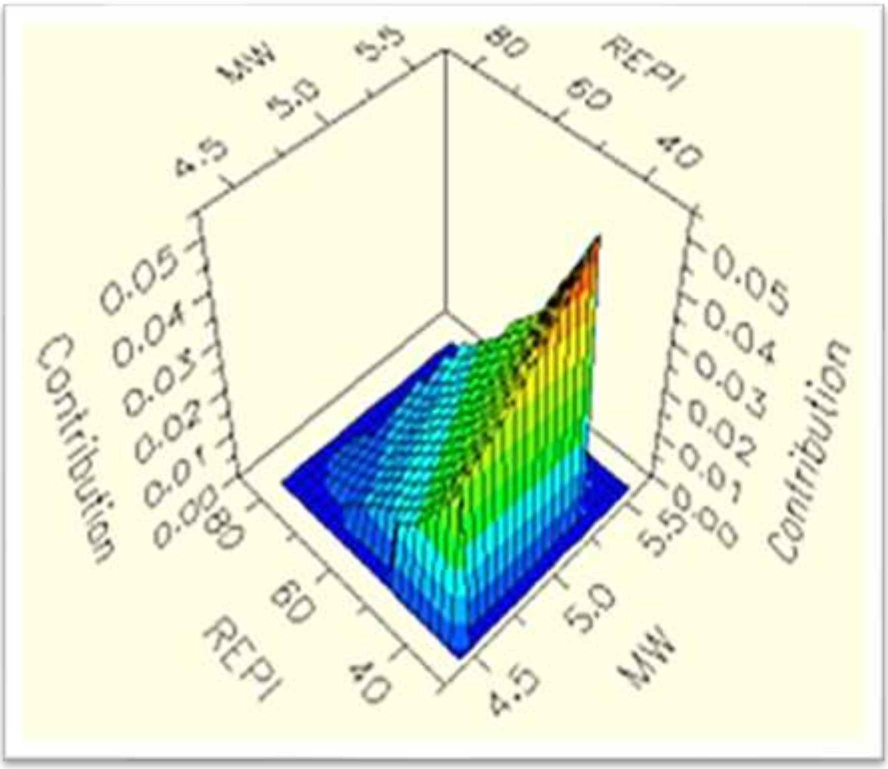

Figure 1: Graph of the Contribution of the Predictor Variable to the Response Variable

Figure 1 is a three-dimensional graph showing that the contribution value to the Response variable increases as the value of the epicenter distance (Repi) decreases. This indicates that the effect of earthquake damage increases as the distance between the epicenter and the surrounding area becomes closer.

It is also clear that the Response variable's contribution increases as the Magnitude (Mw) variable's value rises; hence, an earthquake's destructive power is directly proportional to its magnitude. Using the greatest PGA values, the regions of Andhra Pradesh with the highest potential for earthquake hazards may be determined once the Prediction Analysis findings have been tested and validated.

The PGA value is determined by factors such as the earthquake's magnitude, depth, and distance from the epicenter. Although other variables, such as the state of the location's bedrock, influence earthquake damage, a high PGA value would, in principle, have a large effect. Policymakers may use the findings of the prediction research to establish criteria for infrastructure development in regions prone to earthquakes. This analysis groups the areas with the greatest earthquake susceptibility in Andhra Pradesh.

V. CONCLUSION

Natural hazard assessment and mitigation in tectonic zones has taken a giant leap forward with the merging of data mining and seismic modeling. Despite their usefulness for understanding geophysical processes, classic seismic models aren't always up to snuff when it comes to managing catastrophe risk in real time. To supplement these models, data mining processes large and complicated information, finds trends, and enables predictive analytics—essential for making quick decisions. Together, these factors strengthen early warning systems, facilitate dynamic risk mapping, and guide the development of more robust infrastructure. To top it all off, this integration is multidisciplinary, which means that earth scientists, data analysts, engineers, and politicians can all work together to build a thorough framework for risk assessment. Problems like data heterogeneity, model validation, and ethical concerns, however, need strong frameworks and global collaboration to solve. To fully use this method in the future, it is crucial to invest in data infrastructure, encourage open data sharing, and create hybrid models. In tectonic hazard-prone locations, where climatic change and urbanization are increasing catastrophe risks, safer and more adaptable civilizations may be built by merging data mining with earthquake modeling.

REFERENCES

1. Samm-A, A. S. M. M. Kamal, and M. Z. Rahman, "Earthquake and rainfall-induced landslide hazard assessment of Kutupalong Rohingya camp using meteorological and geological information," *Stochastic Environmental Research and Risk Assessment*, vol. 3, no. 1, pp. 1–4, 2023.
2. D. Priyanto, M. Zarlis, H. Mawengkang, and S. Efendi, "Analysis of earthquake hazards prediction with multivariate adaptive regression splines," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 3, pp. 1–9, 2022.
3. H. D. Nguyen, J. M. LaFave, Y. J. Lee, and M. Shin, "Rapid seismic damage-state assessment of steel moment frames using machine learning," *Engineering Structures*, vol. 252, no. 1, pp. 1–5, 2022.
4. P. S. Nayek and M. Gade, "Artificial neural network-based fully data-driven models for prediction of Newmark sliding displacement of slopes," *Neural Computing and Applications*, vol. 34, no. 11, pp. 6–9, 2022.
5. Sangeeta and B. K. Maheshwari, "Spatial predictive modelling of rainfall- and earthquake-induced landslide susceptibility in the Himalaya region of Uttarakhand, India," *Environmental Earth Sciences*, vol. 81, no. 8, pp. 13–19, Apr. 2022.
6. A. Maqsoom et al., "Efficiency of multiple hybrid techniques for the earthquake physical susceptibility mapping: the case of Abbottabad District, Pakistan," *Environmental Earth Sciences*, vol. 80, no. 19, pp. 3–9, Oct. 2021.
7. Pandit and S. Panda, "Prediction of earthquake magnitude using soft computing techniques: ANN and ANFIS," *Arabian Journal of Geosciences*, vol. 14, no. 13, pp. 3–7, 2021.
8. K. P. Sreejaya, J. Basu, S. T. Raghukanth, and D. Srinagesh, "Prediction of Ground Motion Intensity Measures Using an Artificial Neural Network," *Pure and Applied Geophysics*, vol. 178, no. 6, pp. 2–5, 2021.
9. W. Riyanto, D. Irawan, T. J. Wahyu Adi, D. Iranata, and A. R. Amalia, "Earthquake Vulnerability Assessment of High-Rise Buildings in Surabaya using RViSITS Android Application," in *IOP Conference Series: Materials Science and Engineering*, vol. 739, no. 3, pp. 1–3, 2020.

10. Sangeeta, B. K. Maheshwari, and D. P. Kanungo, "GIS-based pre- and post-earthquake landslide susceptibility zonation with reference to 1999 Chamoli earthquake," *Journal of Earth System Science*, vol. 129, no. 1, pp. 51-55, Dec. 2020.
11. D. Priyanto, M. Zarlis, H. Mawengkang, and S. Efendi, "Approach of Analysis of Data Mining Prediction in Earthquake Case Using Non Parametric Adaptive Regression Method," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 1, no. 4, pp. 247–253, 2020.
12. Merciu, I. Ianos, G.-L. Merciu, R. Jones, and G. Pomeroy, "Mapping accessibility for earthquake hazard response in the historic urban centre of Bucharest," *Natural Hazards and Earth System Sciences*, vol. 18, no. 7, pp. 2011–2026, Jul. 2018.
13. P. H. Nyimbili, T. Erden, and H. Karaman, "Integration of GIS, AHP and TOPSIS for earthquake hazard analysis," *Natural Hazards*, vol. 92, no. 3, pp. 13–17, 2018.
14. N. Rahman, M. A. Ansary, and I. Islam, "GIS-based mapping of vulnerability to earthquake and fire hazard in Dhaka city, Bangladesh," *International Journal of Disaster Risk Reduction*, vol. 13, no. 3, pp. 1–4, 2015.
15. S. Stein, B. D. Spencer, and E. M. Brooks, "Metrics for Assessing Earthquake-Hazard Map Performance," *Bulletin of the Seismological Society of America*, vol. 105, no. 4, pp. 2160–2173, Aug. 2015.
16. W. Zhang, A. Goh, and Y. Zhang, "Multivariate Adaptive Regression Splines Application for Multivariate Geotechnical Problems with Big Data," *Geotechnical and Geological Engineering*, vol. 34, no. 1, pp. 1–8, 2015.
17. F. Yerlikaya-Özkurt, A. Askan, and G.-W. Weber, "An alternative approach to the ground motion prediction problem by a non-parametric adaptive regression method," *Engineering Optimization*, vol. 46, no. 12, pp. 1651–1668, 2014.
18. W. Zhang and A. Goh, "Multivariate adaptive regression splines and neural network models for prediction of pile drivability," *Geoscience Frontiers*, vol. 23, no. 1, pp. 3–7, 2014.
19. H. Güllü, "Prediction of Peak Ground Acceleration by Genetic Expression Programming and Regression: A Comparison Using Likelihood-Based Measure," *Engineering Geology*, vol. 1, no. 4, pp. 92–113, 2012.
20. P. Samui and P. Kurup, "Multivariate adaptive regression spline (MARS) and least squares support vector machine (LSSVM) for OCR prediction," *Soft Computing*, vol. 16, no. 8, pp. 1–8, 2012.
21. Pal et al., "Earthquake hazard zonation of Sikkim Himalaya using a GIS platform," *Natural Hazards*, vol. 45, no. 3, pp. 4–9, 2008.