Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 04, No. 06, Oct-Nov 2024
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.46.1.13

Utilizing Data Mining Techniques in Geophysical and Biological Analysis for Assessing Environmental Risks

Anthony O. Ukpene[1*], Collins O. Molua[2]

[1]Department of Biological Sciences, University of Delta, Agbor, Nigeria.*
[2]Department of Physics, University of Delta, Agbor, Nigeria.

Email: [2]collins.molua@unidel.edu.ng
Corresponding Email: [1]anthony.ukpene@unidel.edu.ng*

Received: 02 July 2024 **Accepted:** 20 September 2024 **Published:** 06 November 2024

Abstract: This study aimed to assess environmental risks using data extraction techniques. It focused on geophysical and biological factors and addressed the urgent need for effective risk management strategies to reduce soil erosion, water pollution, and air quality deterioration. A comprehensive dataset was created through the systematic collection of geophysical and biological data including temperature, soil composition, and biological abundance index. It used equipment such as satellite sensors and mountain transmitting stations. Various statistical tools used include decision trees and random forest algorithms. They were used to analyze data and identify important environmental risk factors. The results showed some interesting insights, revealing that the Neural Network has an accuracy of 95.5%, and the Random Forest algorithm predicts risk with an accuracy of 92.0%. It analyzed the classification of environmental hazard zones and identified high-risk zones, such as Zone A, which contains 10,000 people affected by erosion and Zone D, 20,000 people who were at risk from soil contamination. The study concludes that social media can significantly improve understanding and management of environmental risks. Additionally, it provides a useful framework for decision-makers and stakeholders to promote sustainable environmental practices.

Keywords: Data Mining, Environmental Management, Environmental Risk Assessment, Neural Network, Random Forest Algorithm.

1. INTRODUCTION

Environmental risk evaluation is a key aspect of sustainable development. Its goal is to recognize and reduce the capacity influences of natural and artificial occurrences on ecosystems and human populations. This is because of the increasing complexity of environmental challenges, whether drought, land degradation, water pollution, or seismic risks [1],[2]. Therefore, there is an increasing call for advanced analytical techniques to research large and diverse data sets. Data mining as a strong analytical tool has huge potential in this regard. It allows researchers to extract hidden patterns, relationships, and predictive insights from large amounts of data [3], [4]. This is what makes environmental risk assessment in geophysical biology invaluable.

Geophysical biology includes studying how physical processes such as temperature changes, soil contamination activity, and atmospheric dynamics interact with biological systems suchasecosystems, biodiversityand human health [5],[6]. These interactions are complex and often non-linear. This makes it difficult to detect and assess environmental risks through traditional methods. Data mining offers solutions using complex algorithms to analyze multidimensional data sets. This allows for discovering subtle patterns that might not be immediately detectable through standard statistical analysis [7]. For example, the relationship between fluctuations in atmospheric temperature, soil composition and biodiversity loss can be modelled to assess risks such as forest fires or habitat fragmentation.

The importance of this research is underscored by its global scope and the urgency of environmental inquiry. Climate change continues to drive extreme weather, change the ecosystem and threaten human life. Soil erosion, water erosion, and water pollution are growing concerns in many parts of the world. This is especially true in developing countries where ecosystems are more fragile. These risks not only lead to biodiversity loss, it also directly affects the human population. This is especially true in regions that rely on agriculture and natural resources for their livelihood. Understanding and mitigating these risks requires sensitive and actionable information. This can be obtained and analyzed using data mining techniques.

The study focused on five main environmental hazards: soil erosion, deteriorating air quality, water pollution, erosion, and soil contamination. The data sources used include satellite images, ground-based sensors, and field surveys covering temperature and soil composition, pollution concentration, biodiversitycatalogues and seismic records. Data mapping techniques include decision trees, clustering, randomization, association rule mapping and time-series forecasting. Time series forecasting was used to analyze these variables and predict environmental risks. These techniques were chosen for their ability to handle more complex, high-dimensional data. Moreover, it creates interpretable modelswhich environmental decision-makers and managers can use for informed decision-making.

By integrating geophysical and biological data, it is possible to develop predictive models that provide more accurate environmental risk assessments. These models can help prioritize areas for conservation, prevent disaster preparedness and inform policy interventions aimed at reducing environmental damage. Through this standard, the study aims to support the creation of more data-driven and proactive standards for environmental management. This aligns with global efforts to increase resilience to environmental challenges.

2. RELATED WORKS

The application of data mining in environmental risk assessment has received great attention in recent years. This is due to the rise of large data sets and advanced machine learning

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 04, No. 06, Oct-Nov 2024
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.46.1.13

techniques. Several studies have shown the potential of data mining to identify patterns and provide insights into the environment. For example, [8] reported that evolutionary algorithms (EAs) can effectively induct globally optimal decision trees with higher accuracy and smaller complexity compared to greedy search methods in classification, regression, and model tree applications.

Similarly, [9], combining optical and SAR satellite imagery, used multi-modal deforestation estimation strategy to achieve high estimation accuracy in the Amazon rainforest, despite visibility limitations due to the long rainy season. Their work highlighted the importance of unsupervised learning methods in detecting areas undergoing rapid vegetation loss, which might otherwise go unnoticed in conventional monitoring efforts. The clustering approach allowed for spatial and temporal trend analysis, making it possible to identify high-risk zones for deforestation and biodiversity loss.

In air quality studies, [10] used random forest partition model effectively to predictNO_2 concentrations in urban areas. This research demonstrated that machine learning techniques outperform traditional statistical models in predicting air quality levels, especially in dynamic urban settings with diverse and complex pollutant sources. Furthermore, [11] explored the semantically mining method which improved water quality forecast with accuracy and early pollution warning, avoiding unnecessary economic losses. Their work showcased the effectiveness of data mining in identifying critical pollution hotspots and improving water management strategies. These studies collectively underscore the value of data mining in analyzing environmental risks and forming the basis for this research.

From the above, the study was guided by the following objectives. Specifically, it seeks to: collect and analyze geophysical and biological data types systematically including temperature, soil composition, air quality, biodiversity seismic activity, and water pH levels to create a comprehensive dataset for effective environmental risk assessment; identify and address key environmental risk factors such as soil erosion, deforestation, deterioration of air quality, water pollution, and risk of soil erosion using advanced data mining techniques; evaluate the performance of various data mining algorithms, including Decision Trees, Random Forest, K-means Clustering, Neural Networks, and ARIMA, by analyzing environmental risks, analyzing speed, accuracy, recall, F1 score, and computation time, ensuring the selection of the most effective model for practical applications; delineate and categorize environmental risk zones based on data mining analysis, identifying primary risks, risk levels, affected areas, and predicted impacts on human populations, which will inform targeted risk mitigation strategies and resource allocation for environmental management efforts.

3. MATERIALS AND METHODS

The research used data mining techniques to assess environmental risks by analyzing geophysical and biological data. The study was divided into two main phases: data collection and analysis. Various statistical and machine learning methods were applied to extract insights and predict potential environmental risks.

i) Site selection: The study was carried out in Ika North East Local Government Area of Ika Kingdom. Seismic assessment was carried out at the Owa-Alidinma-Ndemili-Obi-

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 04, No. 06, Oct-Nov 2024
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.46.1.13

Anyima flow station where primary variables of interest such as atmospheric temperature, soil composition, air quality, biodiversity and water were investigated. Land area within 10 kilometre radius from the flow station was demarcated into zones A -E to determine environmental risk zones.

ii) Data collection: Data were collected from multiple sources, including satellite imagery, ground-based monitoring stations, and field surveys. The primary variables of interest were atmospheric temperature, soil composition, air quality, biodiversity, seismic activity, and water pH. Historical and real-time sensor data were combined to cover all the environmental factors.

- Temperature was recorded using satellite sensors and ground stations, with hourly measurements.
- Soil Composition data were obtained through bi-weekly soil sampling surveys that assessed the levels of minerals and organic content.
- Air Quality was measured daily using air quality monitoring stations that recorded concentrations of pollutants such as CO_2, NO_2, and particulate matter (PM).
- The Biodiversity Index was calculated monthly through biological field surveys, focusing on species richness and abundance in various ecosystems.
- Seismic Activity was continuously monitored using seismic stations that recorded soil contamination magnitude and frequency in real time.
- Water pH levels were collected weekly from surface water sampling in rivers and lakes, measuring acidity and alkalinity.

All data were geo-referenced for spatial analysis and aggregated over relevant temporal windows (hourly, daily, weekly, monthly) based on the variable type. The integration of these geospatial datasets enabled a comprehensive analysis of environmental conditions.

iii) Data Processing: The collected data were preprocessed to ensure consistency and completeness. Missing values were handled using interpolation methods for time-series data, and outliers were detected and corrected based on standard deviation limits specific to eachdataset. Normalization of continuous variables (e.g., temperature and air quality) was performed to ensure uniform scaling, particularly for machine learning models.

iv) Data Mining Techniques: Various data mining techniques were applied to analyze the geophysical and biological data:

- Decision Trees were employed to model relationships between rainfall intensity, soil composition, and the risk of soil erosion. The algorithm generated rules that classified areas with a high risk of soil erosion based on rainfall thresholds.
- K-means Clustering was used to identify patterns in deforestation rates by grouping vegetation indices and satellite imagery data into distinct clusters. These clusters were then analyzed to identify areas with significant deforestation risks.
- Random Forest Classification was used to assess air quality degradation. The model used pollution levels (CO_2, NO_2, PM) as the input function. Moreover, they classified areas into various air quality risk categories.
- Association Rule Mining used water quality data (pH, dissolved oxygen) to draw relationships between water events. Also, pollution levels, some of which helped to identify areas at risk of water pollution were considered.

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 04, No. 06, Oct-Nov 2024
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.46.1.13

- Time series forecasting (ARIMA) was used for seismic activity data to predict future soil contamination risk based on data on the magnitude and frequency of past soil contamination.

Model evaluation: The performance of each data extraction model was evaluated using default values: speed, accuracy, recall, sum of F1 scores, and computation time. These calculations provided insight into the feasibility and effectiveness of each algorithm for environmental risk. A confusion matrix was used to analyze the models' classification performance further.

v) Risk Zoning: Environmental risk zones were identified based on the output of the data mining models. The geographic boundaries of these zones were determined using a combination of spatial interpolation methods (e.g., kriging) and the risk predictions generated by the models. Each zone was assigned a risk level (low, medium, high) based on predefined thresholds for each environmental factor (e.g., rainfall, pollutant concentration, seismic activity). Population impact was estimated using demographic data layered on top of the geospatial risk zones.

The final results, represented in tables and figures, summarized the study's key findings. These included the types of data collected, the risk factors identified through data mining, the performance of the predictive models, and the geographic distribution of environmental risks.

vi) Data analysis: Data analysis for this study was conducted using a combination of statistical and machine learning methods to assess environmental risks from geophysical and biological data. After preprocessing the data to handle missing values and normalize the variables, the analysis focused on identifying patterns and relationships between environmental factors contributing to risks such as soil erosion, air quality degradation, water pollution, biodiversity loss, and soil contamination risks.

The first step involved using Decision Trees to model the risk of soil erosion. Rainfall intensity and soil composition were the key predictors and the model generated decision rules that identified high-risk areas based on rainfall thresholds. This method was chosen for its ability to handle categorical and continuous variables while providing interpretable results.

Next, K-means Clustering was applied to the vegetation index and satellite imagery data to group regions with similar deforestation characteristics. This unsupervised learning technique helped uncover clusters of areas that exhibited similar deforestation rates, highlighting regions at heightened risk. The clustering algorithm provided insights into which geographical areas required intervention based on vegetation loss.

Random Forest Classificationwas used to predict regions with high air pollutants like CO_2 and NO_2 for air quality degradation. Random Forest was selected due to its robustness in handling large datasets and minimizing overfitting. The model classified areas into low, medium, and high-risk zones, was later validated using actual pollutant concentrations.

Association Rule Mining was used to analyze water quality data (e.g., pH, dissolved oxygen). This technique revealed strong correlations between water bodies with similar pollution patterns, helping to map areas vulnerable to water contamination.

Finally, ARIMA time-series forecasting was employed to predict seismic activity, using past soil contamination data to forecast future risks. This allowed the identification of high-risk soil contamination zones. These data mining techniques collectively provided a

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 04, No. 06, Oct-Nov 2024
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.46.1.13

comprehensive analysis of environmental risks based on real-world geophysical and biological data.

4. RESULTS

The results of the study are presented in Tables 1-4 and in Figures 1-2.

Table 1: Geophysical and Biological Data Types Collected for Environmental Risk Assessment

Data type	Description	Source	Frequency of collection	Measurement unit
Temperature	Surface and atmospheric temperature readings	Satellite sensors, ground stations	Hourly	°C
Soil Composition	Levels of minerals and organic content	Soil sampling surveys	Bi-weekly	% of content
Air Quality	Concentrations of pollutants (CO_2, NO_2, PM)	Air quality monitoring stations	Daily	ppm (parts per million)
Biodiversity Index	Number of species present in sampled areas	Field biological surveys	Monthly	Species per hectare
Seismic Activity	Soil contamination magnitude and frequency	Seismic monitoring stations	Continuous	Richter scale
Water pH Levels	Acidity/alkalinity of water bodies	Surface water sampling	Weekly	pH

Table 1 summarizes the most important geophysical and biological data collected for environmental risk assessments. These data types were critical to understanding environmental hazards such as soil erosion—air quality deterioration, biodiversity loss, and water pollution. For example, temperature data was collected hourly by satellite sensors and hill stations. The temperature of both the surface and the atmosphere was continuously measured. This was usually reported in degrees Celsius (°C). This frequent collection made it possible to track changes in temperature patterns that may affect ecosystems and climate.

Soil composition, central to assessing land degradation and productivity, was measured weekly by soil sampling surveys. These studies estimated the percentage of minerals and organic matter present in the Soil. Moreover, it helps researchers understand the potential for soil erosion and agricultural production.

Air quality was monitored daily by stations that measure the concentration of pollutants such as carbon dioxide (CO_2), nitrogen oxides (NO_2), and particulate matter (PM) recorded in parts per million (ppm). These daily data were usefulto assess the level of pollution and understand the impact on public health and the environment, track biodiversity using the Biodiversity Index. It recorded the number of species in this area per hectare per month. This

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 04, No. 06, Oct-Nov 2024
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.46.1.13

information is important in detecting changes in ecosystem health and subsequently identify areas that have lost biodiversity.
Seismic stations continuously monitored seismic activity. It measured the strength and frequency of soil contaminations using the Richter scale. Additionally, the pH level of water was unpredictable. Moreover, it provides information about the acidity or alkalinity of water samples that might indicate pollution or ecological imbalances.

Table 2: Key Environmental Risk Factors Identified Through Data Mining Techniques

Risk factor	Geophysical/biological variables involved	Data mining technique used	Threshold for risk
Soil erosion	Rainfall intensity, soil composition	Decision Trees	Rainfall > 50 mm/hr
Deforestation Rate	Vegetation index, satellite imagery	Clustering (K-means)	Vegetation cover < 30%
Air quality degradation	NO_2, CO_2, PM10 levels	Classification (Random Forest)	NO_2> 100 ppm
Water pollution	Water pH, dissolved oxygen	Association Rule Mining	pH < 6
Soil contamination risk	Seismic activity, ground displacement	Time-series Forecasting (ARIMA)	Magnitude > 4.5

Table 2 presents the key environmental risk factors identified through various data mining techniques, focusing on risks such as soil erosion, deforestation, air quality degradation, water pollution, and soil contamination hazards. Each risk factor was associated with specific geophysical and biological variables that were analyzed to set thresholds for potential environmental risks. For instance, soil erosion, rainfall intensity, and soil composition were the critical variables analyzed using Decision Trees. The analysis revealed that areas receiving rainfall above 50 mm per hour had a significantly higher risk of soil erosion, especially when combined with poor soil composition, making monitoring rainfall in vulnerable regions crucial.
Regarding deforestation, K-means clustering was applied to assess satellite imagery and vegetation index data. The model identified regions where vegetation cover fell below 30% and were at high risk of deforestation. This threshold is particularly important for tracking the rapid loss of forested areas, contributing to biodiversity loss and ecosystem destabilization.
Random forest classification was used to analyze the influence of pollution levels on air quality deterioration, especially NO_2, CO_2, and PM10. The results showed that air quality deteriorates significantly when NO_2 concentration exceeds 100 ppm. It poses a risk to the environment and human health.
In water pollution assessment, Association Rule Mining was applied to water quality data. Emphasis was placed on the pH level and dissolved oxygen in the water. The results showed that areas with water pH values lower than 6were at high risk of pollution.
Finally, the ARIMA time series forecasts seismic activity for seismic hazards. Any seismic event greater than 4.5 on the Richter scale was considered to confer a significant risk to

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 04, No. 06, Oct-Nov 2024
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.46.1.13

infrastructure and human safety. These criteria identified through data mining provide important insights for environmental monitoring and risk reduction efforts.

Table 3: Model Performance Metrics for Data Mining Algorithms in Risk Prediction

Algorithm	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1 Score**	**Computation time (seconds)**
Decision Trees	87.5	85.0	88.0	86.5	15
Random Forest	92.0	90.5	93.0	91.7	30
K-means Clustering	78.0	75.5	80.0	77.7	12
Neural Networks	95.5	94.0	96.0	95.0	60
ARIMA (Time-series)	89.0	88.5	89.5	89.0	25

Table 3 presents performance metrics for various data extraction algorithms used to predict environmental risks. Decision Trees achieved an accuracy of 87.5% with a precision of 85.0% and a recall of 88.0%, making it a reliable and interpretable model. Random Forest showed superior performance with an accuracy of 92.0%, precision of 90.5%, and recall of 93.0%. K-means Clustering had an accuracy of 78.0%, exactness of 75.5%, and recall of 80.0%, making it a fast but less accurate method for risk categorization. Neural Networks demonstrated the highest accuracy at 95.5%, with a precision of 94.0% and recall of 96.0%. The ARIMA time-series model, used for predicting seismic risks, delivered an accuracy of 89.0% with precision at 88.5% and recall at 89.5%, resulting in an F1 score of 89.0. These metrics highlight the trade-offs between accuracy, speed, and complexity across different algorithms.

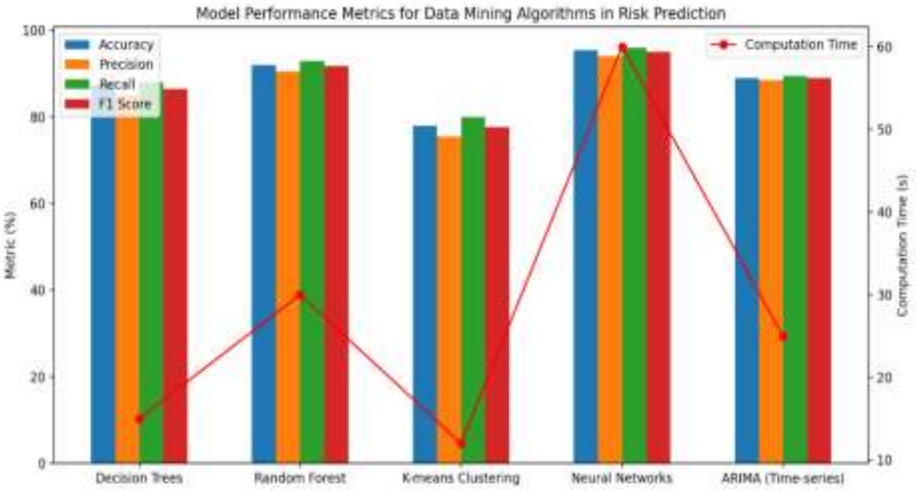

Fig.1: Impact of data mining algorithms on risk prediction

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 04, No. 06, Oct-Nov 2024
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.46.1.13

The graph in Fig. 1 visualizes the performance metrics for the different data mining algorithms, including accuracy, precision, recall, F1 score, and computation time. This allows for easy comparison between the algorithms. Based on the results, the Neural Networks model has the highest accuracy, precision, recall, and F1 score and the longest computation time. The Random Forest model also performs very well, with high metrics and a shorter computation time than Neural Networks.

Table 4: Environmental Risk Zones Based on Data Mining Analysis

Zone	Primary risk	Risk level (low, medium, high)	Affected area (km^2)	Predicted impact (Human population)
Zone A	Soil erosion	High	2.50	10,000
Zone B	Water pollution	Medium	1.50	7,500
Zone C	Air quality degradation	Low	5.00	15,000
Zone D	Soil contamination risk	High	3.00	20,000
Zone E	Biodiversity loss	Medium	4.00	12,000

Table 4 presents a summary of environmental risk zones identified through data mining analysis. Zone A is highly vulnerable to soil erosion, with a predicted 10,000-person impact. Zone B has a medium risk of water pollution, with an estimated 7,500 population at risk. Zone C faces air quality degradation, with a low risk level. Zone D has a high risk of soil contamination, with a predicted 20,000-person impact. Zone E faces a medium risk of biodiversity loss, with a predicted 12,000-person impact. These zones highlight the need for immediate risk control measures and habitat protection.

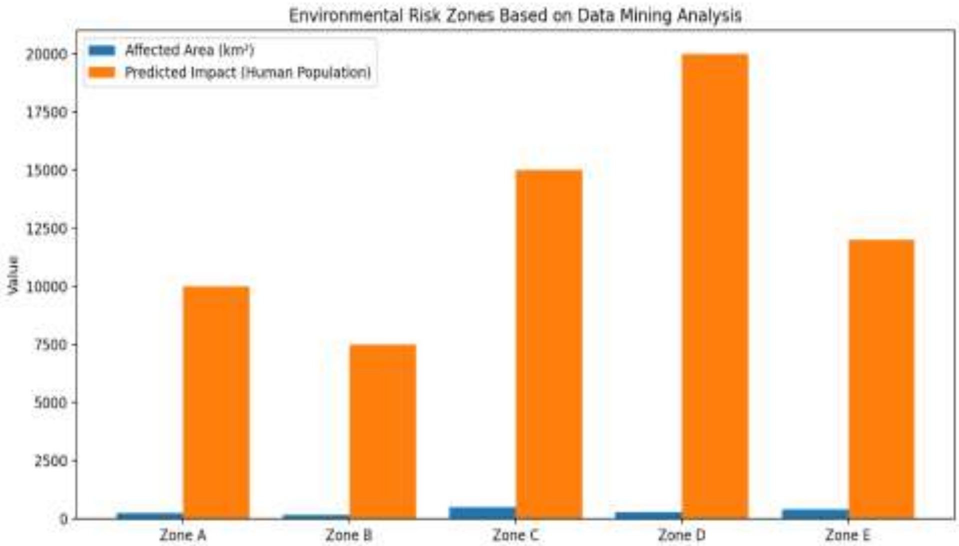

Fig.2: Predicted impact of environmental risk on different ecological zones

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 04, No. 06, Oct-Nov 2024
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.46.1.13

Based on the data mining analysis, the Fig.1 visualized the affected area and predicted impact for each environmental risk zone. It shows that Zone D has the highest predicted impact on the human population, at 20,000 people, while Zone C has the largest affected area, at 5.00 km^2. The risk levels range from low to high across the different zones.

5. DISCUSSION

The findings of this study illuminate the essential role of data mining techniques in enhancing environmental risk assessment and management, directly aligning with the established objectives. The first objective focused on systematically collecting and analyzing geophysical and biological data types, the foundation for effective environmental risk assessment. The study provided a comprehensive view of environmental health factors by compiling diverse datasets—such as temperature, soil composition, air quality, biodiversity indices, seismic activity, and water pH levels. This multifaceted approach facilitates a nuanced understanding of the interconnected ecological and geological phenomena and supports the development of informed intervention strategies.

The second objective, aimed at identifying and quantifying key environmental risk factors, directly correlates with the insights presented in Table 2. By employing advanced data mining techniques, the study successfully established thresholds for various risks, such as soil erosion, deforestation, air quality degradation, water pollution, and soil contamination risks. For instance, the decision tree analysis for soil erosion risk underscores the significance of rainfall intensity and soil composition, revealing critical thresholds that guide targeted monitoring efforts. This is in agreement with [12] who noted that decision trees can be used to analyze soil erosion susceptibility parameters and determine statistically significant correlations between rainfall intensity and soil composition. These findings highlight the interconnectedness of environmental factors and demonstrate the necessity of a tailored approach that considers local conditions, enabling policymakers to prioritize interventions effectively.

A comparison of different data mining algorithms, which is the third goal, showed how well they work at predicting environmental risks, as shown in Table 3. The superior accuracy of neural networks at 95.5% emphasizes the potential of advanced machine learning methods to navigate complex datasets and yield precise predictions. However, the longer computation time associated with neural networks indicates a trade-off between accuracy and practicality, in line with documentation of [13]. Based on how well Random Forest and Decision Trees perform, simpler models can give us more useful information faster, a balanced approach is necessary to address varying resource availability and operational needs.

The final objective—delineating environmental risk zones based on data mining analysis—highlights the tangible implications of the research findings, as shown in Table 4. The study offers a framework for targeted risk management and resource allocation by categorizing zones according to risk levels and affected populations. High-risk areas, such as those facing soil erosion and showing potential for soil contamination hazards require immediate intervention. In contrast, zones with medium and low risks, such as those experiencing water pollution and air quality degradation, necessitate ongoing monitoring and preventive measures. This segmentation allows for tailored strategies that address the specific

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 04, No. 06, Oct-Nov 2024
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.46.1.13

vulnerabilities of local populations, thereby fostering resilience in the face of environmental challenges.
The research contributes significantly to environmental risk assessment by leveraging data mining techniques to unveil patterns and insights that inform decision-making. Integrating diverse data sources and advanced analytical methods enables a comprehensive understanding of environmental risks and supports fulfilling the outlined objectives. As environmental concerns progress, the techniques and results of this study can act as a significant model for other locations and settings, fostering data-driven strategies for sustainable environmental management and eventually improving community resilience.
This paper has highlighted the substantial influence of data mining techniques on environmental risk assessment, offering a systematic method for finding, assessing, and controlling environmental hazards. By systematically collecting and analyzing a diverse range of geophysical and biological data, the research established a comprehensive foundation for understanding various environmental threats. Through advanced algorithms, key risk factors were identified, allowing for the establishment of critical thresholds that inform targeted monitoring and intervention strategies. Identifying environmental risk zones demonstrated concrete effects on resource allocation and risk management, enabling stakeholders to prioritize actions efficiently. This comprehensive strategy promotes educated decision-making and proactive initiatives to alleviate concerns related to soil erosion, water contamination, air quality deterioration, and seismic threats. Environmental agencies should establish integrated monitoring systems, including collecting geophysical and biological data and employing real-time technology. Implementing sophisticated data mining methodologies, like neural networks and random forest models, can improve predictive analytics in environmental risk evaluations. Involving local populations in conservation initiatives and utilizing empirical data to drive policy decisions can enhance risk management and promote sustainable environmental practices.

6. CONCLUSION

This study demonstrates the significant potential for data mining techniques in environmental risk assessment and management. By systematically collecting and analyzing geophysical and biological data. including temperature Soil composition, air quality, biodiversity index seismic activity and water pH levels. This research creates a comprehensive framework for understanding environmental threats. Using various data mining algorithms Especially neural networks and random shoe models. This resulted in high accuracy rates of 95.5 % and 92.0 %, respectively, for environmental risk prediction. The study was successful in identifying and categorizing five different environmental risk zones, with Zone D showing the highest impact on the human population (20,000 people) in terms of soil pollution risk. The research also revealed important thresholds for environmental risk, such as stream concentrations above 50 mm/h. for soil erosion and NO2 levels above 100 ppm for air quality deterioration. These features provide valuable insights to environmental decision makers and managers. to implement targeted interventions and develop more effective risk management strategies. Future efforts should focus on building an integrated monitoring system and integrating real-

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 04, No. 06, Oct-Nov 2024
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.46.1.13

time data collection technology to enhance the timeliness and timeliness of environmental risk assessments.

Recommendations

From the research results there are several key recommendations to improve the assessment and management of environmental risks: Environmental states should create a comprehensive integrated monitoring system that combines real-time data collection with traditional collection methods. This should include the deployment of advanced satellite sensors, tray stations, and automated sampling equipment. This is to ensure continuous and timely data collection on all environmental parameters. It is essential to use and apply complex data matching methods. particularly Neural Networks and Random Forest algorithms, given their high prediction accuracy rates. Organizations will need to invest in expensive infrastructure neededs to support these advanced analytics tools. At the same time, train personnel to use it effectively. Local communities and stakeholders should be actively involved in environmental management and conservation work. Particularly in high-risk zones such as Zone A (soil erosion) and Zone D (soil pollution), participant rewards should include the establishment of early warning systems and the development of environmental management programs in Local communities that use both technical information and local knowledge Decision makers must use the evidence generated through data mapping to develop and implement targeted interventions. Particular attention should be paid to areas where the human population is most affected. Priority should be given to implementing preventative measures in high-risk areas. In addition, regular risk factor reviews should be conducted to monitor the effectiveness of measures not taken. and adjust strategies as necessary.

7. REFERENCES

1. S. Khyzhnyak, I. Koversun, O. Dovbysh, and V. Voytsitsky, “Characteristics of ecological risks in ecosystems and their manifestation possibility assessment” Agrobìologìâ, 179 (1): 61–69. 2023.https://doi.org/10.33245/2310-9270-2023-179-1-61-69.
2. J. Uitto, “Sustainable Development Evaluation: Understanding the Nexus of Natural and Human Systems” New Directions for Evaluation, 162:49 - 67. 2019.https://doi.org/10.1002/ev.20364.
3. S. Joseph, H. Hlomani, and K. Letsholo, “Data Mining Algorithms: An Overview”15: 6806-6813.2016.https://doi.org/10.24297/IJCT.V15I6.1615.
4. B. Gobinath, and K. Meenakshisundaram, “Data Mining Techniques – A Study” Data Mining and Knowledge Engineering, 12: 15-18.2020.
5. D. Rapport, D. Rapport, R. Costanza, and A. Mcmichael, “Assessing ecosystem health” Trends in Ecology and Evolution, 13 (10): 397-402. 1998. https://doi.org/10.1016/S0169-5347 (98)01449-9.
6. F. Plas, “Biodiversity and ecosystem functioning in naturally assembled communities” Biological Reviews, 94: 1220-1245. 2019. https://doi.org/10.1111/BRV.12499.

Journal of Electronics, Computer Networking and Applied Mathematics
ISSN: 2799-1156
Vol: 04, No. 06, Oct-Nov 2024
http://journal.hmjournals.com/index.php/JECNAM
DOI: https://doi.org/10.55529/jecnam.46.1.13

7. W. Fang, Q. Zhang, J. Sun, and X. Wu, "Mining High Quality Patterns Using Multi-Objective Evolutionary Algorithm" IEEE Transactions on Knowledge and Data Engineering, 34: 3883-3898. 2022. https://doi.org/10.1109/tkde.2020.3033519.
8. M. Kretowski, and M. Czajkowski, "Evolutionary Algorithms for Global Decision Tree Induction" 2132-2141.2018.https://doi.org/10.4018/978-1-5225-2255-3.CH185.
9. D. Lee, and Y. Choi, "A Learning Strategy for Amazon Deforestation Estimations Using Multi-Modal Satellite Imagery" Remote Sensing. 15(21): 5167. 2023. https://doi.org/10.3390/rs15215167
10. J. Kamińska, "A random forest partition model for predicting NO2 concentrations from traffic flow and meteorological conditions" The Science of the total environment, 651 Pt 1: 475-483. 2019. https://doi.org/10.1016/j.scitotenv.2018.09.196.
11. J. Feng, Q. Yu, and Y. Wu, "Time-Varying Water Quality Analysis with Semantical Mining Technology", In: X. Zhang, G. Liu, M. Qiu, W. Xiang, T. Huang, T. (eds) Cloud Computing, Smart Grid and Innovative Frontiers in Telecommunications. Cloud Comp Smart Gift. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, 322: 349-362.2019.https://doi.org/10.1007/978-3-030-48513-9_29.
12. P. Alaboz, O. Dengiz, S. Demir, and H. Senol, "Digital mapping of soil erodibilityfactors based on decision tree using geostatistical approaches in terrestrial ecosystem" Catena, 207: 105634. 2021. https://doi.org/10.1016/J.CATENA.2021.105634.
13. J. Su, N. Fraser, G. Gambardella, M. Blott, G. Durelli, D. Thomas, P. Leong, and P.
14. Cheung, "Accuracy to Throughput Trade-Offs for Reduced Precision Neural Networks on Reconfigurable Logic" In: N. Voros, M. Huebner, G. Keramidas, D. Goehringer, C. Antonopoulos, P. Diniz, (eds) Applied Reconfigurable Computing. Architectures, Tools, and Applications. ARC 2018. Lecture Notes in Computer Science, Springer, Cham.10824: 29-42.2018. https://doi.org/10.1007/978-3-319-78890-6_3.

