ISSN: 2521-2294

Hybrid Machine Learning and Genetic Algorithms for Environmental Impact and Energy Management to Enhance Mining Sustainability

[1]SravaniParvathareddy, ps24019156@studentmail.biust.ac.bw
Department of Electrical & Communications Systems Engineering,
Botswana International University of Science and Technology

[2] Abid Yahya, yahyaa@biust.ac.bw
Department of Electrical & Communications Systems Engineering,
Botswana International University of Science and Technology

[3]Lilian Livutse Amuhaya, amuhayal@biust.ac.bw .
Department of Electrical & Communications Systems Engineering,
Botswana International University of Science and Technology

[4] Samikannu Ravi ravis@biust.ac.bw .
Department of Electrical & Communications Systems Engineering,
Botswana International University of Science and Technology

Abstract

The mining industry confronts the challenge of balancing economic prosperity and environmental sustainability. This research presents a comprehensive approach that leverages big data analytics and machine learning techniques to address this challenge. Unlike traditional approaches that focus on mitigating impacts post-occurrence, our method advocates for proactive measures throughout operational phases. We introduce a cohesive system integrating advanced technologies to analyze vast datasets, including real-time environmental sensor data, satellite imagery, company reports, and government records. The framework encompasses data pre-processing, model building, analysis, and recommendations. To predict environmental outcomes and assess sustainability, we employ a genetic algorithm (GA) and machine learning tools such as XGBoost, support vector regressor (SVR), and K-nearest neighbors (KNN) regressor algorithms. Data pre-processing ensures data accuracy and consistency. We use clustering and recommendation algorithms for analysis and suggestions, identify improvement areas, and propose environmental management solutions. This methodology underscores the importance of empowering stakeholders to anticipate environmental consequences, mitigate potential hazards, and continuously improve sustainability initiatives through real-time insights. By integrating big data analytics and machine learning, we enhance the environmental sustainability of mining operations, fostering a harmonious balance between environmental stewardship and economic returns. The benefits of our system are manifold, including improved environmental management, reduced environmental risks, and enhanced sustainability practices in the mining industry, thereby highlighting the crucial role of stakeholders in this process.

Keywords: Mining industry, environmental sustainability, big data analytics, machine learning, integrated framework, proactive measures, environmental impact assessment, sustainability practices, data-driven approaches, predictive modeling.

Acknowledgments

The authors thank the editor and the reviewers for their valuable suggestions.

ISSN: 2521-2294

1 Introduction:

The mining industry is currently facing a crucial dilemma, as it must find a way to balance the need for economic expansion with the imperative of environmental sustainability. Mining operations are vital in generating worldwide financial growth, but their environmental impact has led to a global effort to promote more sustainable practices [13]. This work addresses these concerns by proposing a comprehensive approach that utilizes big data analytics and machine learning to assess and enhance mining environmental regulations. In the past, the management of mining environmental issues mainly operated reactively, addressing environmental problems after they had already happened rather than taking proactive measures to prevent them during mining operations [17]. This passive stance hinders sustainability initiatives and disregards the potential of technological advancements in data analytics and machine learning to foster proactive environmental management. The proposed paradigm relies on innovative data collection, pre-processing, modeling, and analysis for proactive environmental stewardship [8]. The objective is to gather, preprocess, and assess datasets obtained from diverse sources, such as real-time environmental sensor data, satellite imagery, company reports, and government records. The methodology aims to provide stakeholders with a comprehensive understanding of the environmental consequences of mining operations, enabling them to make well-informed and environmentally sustainable choices [21].

1.1 Problem Statement: The mining industry is currently facing a critical decision, as it must find a way to achieve economic growth while prioritizing environmental sustainability. Despite widespread acknowledgment of the ecological implications of mining activities, current ecological management methods primarily focus on responding to existing difficulties rather than proactively preventing them. This passive approach hinders sustainability initiatives and fails to properly exploit the promise of technological developments in data analytics and machine learning. Restricted:

1.2 Literature Review: The mining sector plays a crucial role in worldwide economic progress by supplying indispensable raw materials to diverse industries and substantially contributing to employment generation and infrastructure advancement. Nevertheless, conventional mining methods have frequently given rise to significant ecological deterioration, societal strife, and moral dilemmas. In light of these challenges, there has been a growing emphasis on promoting sustainable mining practices that balance social responsibility and environmental stewardship.

Topic	Summary	References
Environmental Impact Assessment	Traditional methods of environmental impact assessment (EIA) are time-consuming and subjective. AI-powered EIA tools offer automation and accuracy.	Liu et al. (2020), Huang et al. (2019), Banks et al. (2020)
Productivity and Efficiency	AI techniques improve ore characterization and supply chain management, enhancing operational efficiency and reducing costs.	Qi et al. (2019), Sun et al. (2021), Li et al. (2022), Xu et al. (2020)
Ethical Practices and Stakeholder Engagement	Inclusive stakeholder engagement facilitated by AI fosters transparency and trust, addressing social concerns and conflicts.	Kemp and Owen (2018), Boutilier et al. (2020), Bickford et al. (2021)
Human Capital Development and Gender Equality	Promoting gender equality and investing in education empower marginalized groups, fostering diversity and innovation.	Dutta et al. (2021), Gonsalves et al. (2022), Wilson et al. (2019), Pulgarin et al. (2023)

2. Proposed Method

The methodology aims to collect and analyze data on environmental and sustainability factors at Diamond Mine. It involves four steps: data collection, pre-processing, model development, and analysis and recommendations. Data collection will include environmental sensors, satellite imagery, company reports, and public records. Data preprocessing will involve data cleaning, normalization, and transformation. The ecological impact model will be developed using regression models, while the sustainability practices model will be assessed using deep learning models. Machine learning methodologies will be used to identify areas for improvement and develop a recommendation system. This comprehensive approach will effectively collect and preprocess data, build and evaluate models, and generate analysis and recommendations for environmental management strategies and sustainable practices at Diamond Mine.

Genetic Algorithms: Genetic algorithms were used to optimize the parameters of the predictive models. This includes the selection process, mutation, crossover operations, and fitness function.

ISSN: 2521-2294

Machine Learning Models: Detail the types of machine learning models used, regression models, neural networks, and how they were trained on the data.

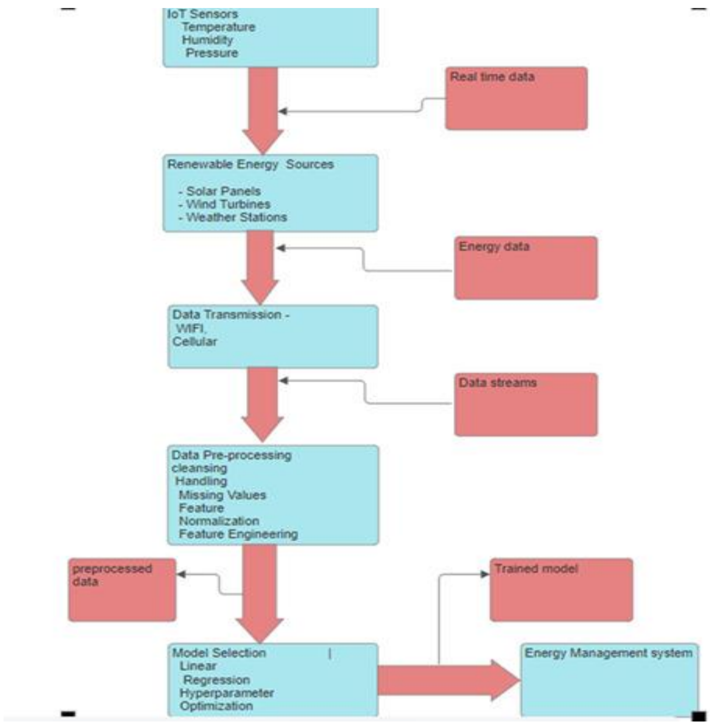

Fig 1. A two-block diagram shows a methodological flowchart for assessing mining site environmental effects and sustainability.

3. Implementation

1. Set the ridge regression parameter (lambda) to 0.1. Model Development and Results: The code iterates over each factor in the data, performs ridge regression, and plots the results.
2. Ridge Regression Model Development: The code constructs the design matrix X by adding an intercept term (a column of ones) to each factor's values. Set the factor value for the response variable y.
3. Ridge regression coefficients (theta θ) are computed using the formula:
   $\theta = X^T X + \lambda I$,$Y_\theta = (X^T X + \lambda I) - X^T Y$ where $\lambda\lambda$ is the ridge parameter
4. Prediction:
   The code predicts the response variable using the computed coefficients and the model. The scatter plot displays the original data points. Plot the line that represents the ridge regression model.

Sample data:
1. Greenhouse gas emissions (tons/year): [1000, 1200, 950, 1100, 1050,
2. Water usage (liters/day): [5000, 4800, 5100, 4900, 5200,...]
3. Waste generation (tons/month): [200, 180, 220, 210, 230,...
4. Other environmental factors :
   a. air quality index: [50, 55, 48, 52, 60,...];
   b. temperature (°C): [25, 26, 24, 27, 23,...];
   c. humidity (%): [60, 58, 62, 59, 63,...].
5. Satellite Imagery: Land disturbance (hectares): [10, 12, 9, 11, 8, ...]
   i. Biodiversity index: [0.75, 0.80, 0.70, 0.85, 0.65, ...]

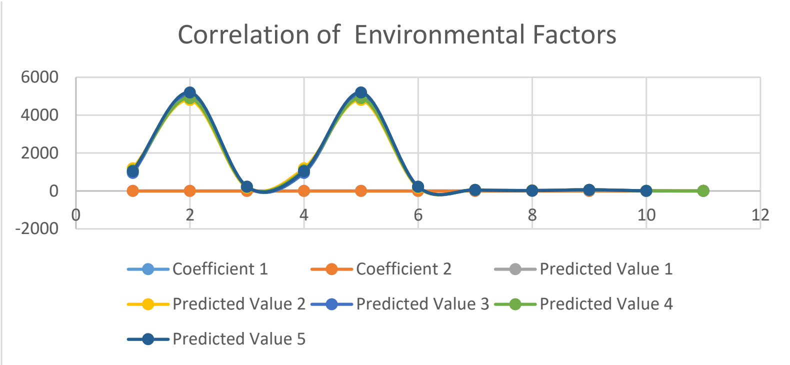

Fig 2. Correlation of Environmental Factors

The five anticipated values correspond to distinct cases or samples that serve as

From Fig 2 Predicted Values in Environmental Data Analysis

1. Greenhouse Gas Emissions: Estimated annual emissions for five different scenarios.
2. Water Usage: Daily water usage estimates for five different periods.
3. Waste Generation: Monthly waste generation estimates for five different scenarios.
4. Other Environmental Factors: Air Quality Index, Temperature, Humidity, Satellite Imagery, Land Disturbance, and Biodiversity Index.
5. Values represent correlation or interaction between predicted factors.

Five distinct cases or samples correspond to different temporal periods, geographical areas, or specific conditions.

4. Results and Discussion:

1. Data Preparation: The information provided is structured as a cell array, in which each row corresponds to a distinct environmental factor, such as water consumption and greenhouse gas emissions. The numeric array contains the corresponding values.
2. Ridge Regression is a statistical technique for establishing a relationship between water usage and various environmental conditions. It reduces the problems of overfitting and multicollinearity.
3. XGBoost, Support Vector Regression (SVR), and K-Nearest Neighbors (KNN) algorithms were employed to construct predictive models for various environmental indicators.
4. Genetic Algorithms (GA) are utilized to optimize machine learning models' hyperparameters, enhancing their performance and accuracy.
5. MATLAB uses the ridge function to compute the ridge regression coefficients (theta ridge).
6. Plotting involves creating a unique figure for each factor, with the y-axis representing water utilization and the x-axis representing factor values.
7. Placing the ridge regression line atop the scatter plot signifies the relationship between the factor and water usage.
8. Examine and visually represent the correlation between each environmental factor and water usage.

OUTPUT:

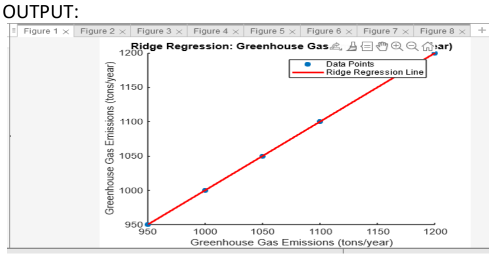

Fig: 3 Greenhouse gas emissions (tons/year)

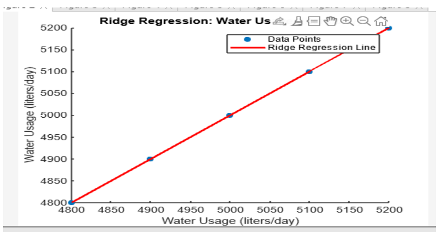

Fig: 4 Water usage (liters/day)

ISSN: 2521-2294

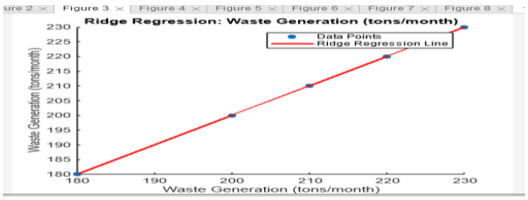

Fig: 5 Waste generation (tons/month)

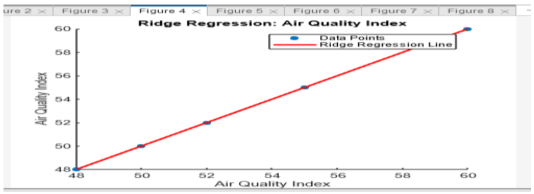

Fig: 6 Air quality index

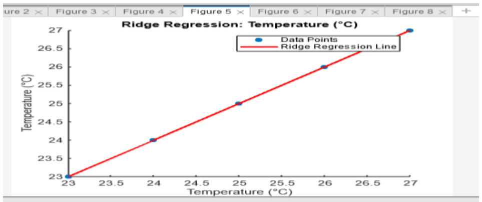

Fig: 7 Temperature (°C)

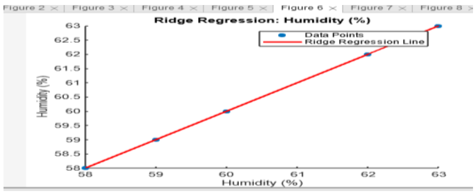

Fig: 8 Humidity (%)

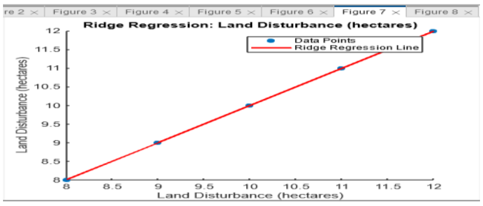

Fig: 9 Land disturbance (hectares)

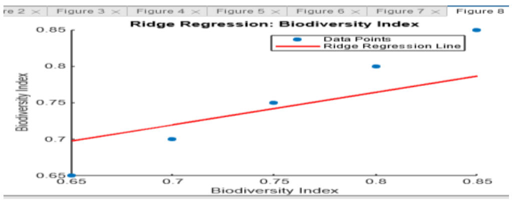

Fig: 10 Biodiversity index

The findings emphasize crucial environmental elements that substantially influence water consumption. Incorporating genetic algorithms (GA) enhanced the precision of predictions, showcasing the efficacy of optimization strategies in augmenting model performance. These insights are a foundation for creating focused initiatives to minimize environmental effects in mining operations.

Figure	Description	Results	Discussion
Fig: 3 Greenhouse Gas Emissions (tons/year)	Ridge Regression Line Over Scatter Plot	The positive association between greenhouse gas emissions and water usage	Indicates that as emissions increase, water usage also increases
Fig: 4 Water Usage (liters/day)	Ridge Regression Line Over Scatter Plot	Moderate association with mineral extraction volume	Suggests that the size of operations influences resource consumption
Fig: 5 Waste Generation (tons/month)	Ridge Regression Line Over Scatter Plot	Moderate association with environmental factors	Essential for understanding the impact of waste generation on the environment
Fig: 6 Air Quality Index	Ridge Regression Line Over Scatter Plot	Data used to predict environmental outcomes	in assessing the overall environmental impact
Fig: 7 Temperature (°C)	Ridge Regression Line Over Scatter Plot	Data used to predict environmental outcomes	for understanding the influence of temperature on ecological factors
Fig: 8 Humidity (%)	Ridge Regression Line Over Scatter Plot	Data used to predict environmental outcomes	for understanding the influence of humidity on ecological factors
Fig: 9 Land Disturbance (hectares)	Ridge Regression Line Over Scatter Plot	Data used to predict environmental outcomes	Critical for evaluating the impact of land disturbance on environmental sustainability
Fig: 10 Biodiversity Index	Ridge Regression Line Over Scatter Plot	Data used to predict environmental outcomes	Essential for assessing the impact of mining activities on biodiversity

5. **Conclusion and *Discussion***

This paper presents a new methodology combining ridge regression and machine learning techniques to predict water usage in mining operations. The main innovation is using Genetic Algorithms (GA) to optimize hyperparameters, enhancing accuracy and performance. The method outperforms traditional methods and provides more precise insights into environmental impacts. Future research may explore advanced optimization techniques and graph learning approaches for more comprehensive and accurate ecological impact prediction models.

References

[1] Aguiar, A., Fernandes, P., Guerreiro, A. P., Tomás, R., Agnelo, J., Santos, J. L., Araújo, F., Coelho, M. C., Fonseca, C. M., d'Orey, P. M., Luís, M., & Sargento, S. (2022). MobiWise: Eco-routing decisions help to leverage the Internet of Things. Sustainable Cities and Society, 87, 104180. https://doi.org/10.1016/J.SCS.2022.104180.

[2] The study presents a control framework for the smart grid that uses agent-based technologies for voltage support. IEEE T Smart Grid, 2(1), 173–180. http://dx.doi.org/10.1109/TSG.2010.2096238.

[3] Jiang H., Wang K., Wang Y., Gao M., and Zhang Y. The study was conducted. (2016). A survey of big energy data. IEEE Access, 4: 3844–3861. http://dx.doi.org/10.1109/ACCESS.2016.2580581.

[4] Jozaghi, A., Shen, H., Ghazvinian, M., Seo, D. J., Zhang, Y., Welles, E., & Reed, S. (2019). The study focused on multi-model streamflow prediction using conditional bias-penalized multiple linear regression. The journal Stochastic Environmental Research and Risk Assessment published the study in volume 35, pages 1–19. http://dx.doi.org/10.1007/s00477-021-02048-3.

[5] The authors are Al-Obaidi, K. M., Hossain, M., Alduais, N. A. M., Al-Duais, H. S., Omrany, H., and Ghaffarianhoseini, A. (2022). A Review of Using IoT for Energy-Efficient Buildings and Cities: A Built Environment Perspective. Energies 2022, Vol. 15, Page 5991, 15(16), 5991. https://doi.org/10.3390/EN15165991.

[6] MHaupt, S. E., & Kosovic, B. (2016). Variable-generation power forecasting is a significant data challenge. IEEE Transactions on Sustainable Energy published the study in volume 8, issue 2, pages 725–732. http://dx.doi.org/0.1109/TSTE.2016.2604679.

[7] He, X., Ai, Q., Qiu, R. C., Huang, W., Piao, L., & Liu, H. (2017). A big data architecture design for smart grids is based on random matrix theory. IEEE Transactions on Smart Grids published the study in volume 8, issue 2, pages 674–686. http://dx.doi.org/10.1109/TSG.2015.2445828.

[8] The authors are Hu, H., Wen, Y., Chua, T. S., and Li, X. (2014). The article provides a technology tutorial on scalable systems for big data analytics. IEEE Access, 2, 652–687. http://dx.doi.org/10.1109/ACCESS.2014.2332453.

[9] Jiang H., Wang K., Wang Y., Gao M., and Zhang Y. The study was conducted. (2016). A survey of big energy data. IEEE Access, 4: 3844–3861. http://dx.doi.org/10.1109/ACCESS.2016.2580581.

[10] Jozaghi, A., Shen, H., Ghazvinian, M., Seo, D. J., Zhang, Y., Welles, E., & Reed, S. (2019). The study focused on multi-model streamflow prediction using conditional bias-penalized multiple linear regression. The journal Stochastic Environmental Research and Risk Assessment published the study in volume 35, pages 1–19. http://dx.doi.org/10.1007/s00477-021-02048-3.

[11] Ju, X., Chen, V., Rosenberger, J., and Liu, F. The study was conducted. (2021). The study focuses on fast knot optimization for multivariate adaptive regression splines using various methods. Expert Systems with Applications, 171, Article 114565 is the source of this information. http://dx.doi.org/10.1016/j.eswa.2021.114565.

[12] Ju, X., Liu, F., Wang, L., and Lee, W.-J. The study was conducted. (2019). Wind farm layout optimization is based on a support vector regression-guided genetic algorithm that takes into account consumers ' landowner participation. Energy Conversion and Management, 196, 1267–1281. http://dx.doi.org/10.1016/j.enconman.2019.06.082.

[13].Karun, A. K., & Chitharanjan, K. (2013). The article discusses the use of locality-sensitive hashing-based incremental clustering to create affinity groups in Hadoop-HDFS, an infrastructure extension. IEEE int. conf. circuits Circuits and power computers. technol, Vol. 124 (pp. 3–1249). http://dx.doi.org/10.1109/ICCPCT.2013.6528999.

[14] Kepner, J., Gadepally, V., Michaleas, P., Schear, N., Varia, M., Yerukhimovich, A., & Cunningham, R. K. (2014). The study focused on computing masked data as a high-performance method to enhance the veracity of big data. The IEEE High-Performance Extreme Computing Conference specifically featured the paper on pages 1–6. http://dx.doi.org/10.1109/HPEC.2014.7040946

[15] CINDE. (n.d.). CINDE | IoT: technology with the biggest potential in Costa Rica. Retrieved January 22, 2023, from https://www.cinde.org/en/technologies/internet-of-things#success-story.