ISSN 2085-4552

Comparison of Multilinear Regression and AdaBoost Regression Algorithms in Predicting Corrosion Inhibition Efficiency Using Pyridazine Compounds

Yudha Mulyana[1], Muhamad Akrom[2], Gustina Alfa Trisnapradika[3], Nabila Putri Setiawan[4]
[1,2,3,4]Informatics Engineering Program, Dian Nuswantoro University, Semarang, Indonesia
[1]yudha.myn@gmail.com,[2] m.akrom@dsn.dinus.ac.id, [3] gustina.alfa@gmail.com,[4] nabilasei@gmail.com

Accepted 25 October 2024
Approved 08 January 2025

***Abstract*— Corrosion is a severe problem in various industries, leading to increased production costs, maintenance, and decreased equipment efficiency. Using organic compounds as corrosion inhibitors has become an increasingly desirable solution due to their effectiveness and environmental friendliness. This study compares the performance of two machine learning algorithms, Multilinear Regression (MLR) and AdaBoost Regression (ABR), in predicting the corrosion inhibition efficiency (CIE) of pyridazine-derived compounds. The dataset consists of molecular properties as independent variables and CIE values as targets. A k-fold cross-validation using the k=10 process was used to measure the model's performance, where the dataset was divided into equal subsets. Each iteration uses one subset as validation data and the other as training data. Results show that the AdaBoost Regression model achieves higher accuracy (99%) than Multilinear Regression (98%) in predicting CIE. Important feature analysis showed that Total Energy (TE) and Dipole Moment (μ) were the most influential variables in the ABR model, highlighting their important role in inhibitor effectiveness. Model evaluation was performed with R2 and RMSE metrics, where nonlinear models such as ABR were shown to be superior in predicting corrosion inhibition efficiency. These findings support using nonlinear methods to improve the effectiveness of protecting industrial equipment from corrosion.**

***Index Terms*— Pyridazine; Corrosion inhibitors; AdaBoost Regression; Multilinear Regression; Corrosion inhibition efficiency**.

I. INTRODUCTION

Corrosion is a significant phenomenon in industry that can result in large economic losses, increased production costs, and decreased efficiency and equipment life [1]. The main corrosion causes include air humidity, metal contamination with acidic compounds, sulfur, and other corrosive gases. According to the definition of the International Union of Pure and Applied Chemistry (IUPAC), corrosion is a natural process that occurs in materials such as metals, ceramics, and polymers [2]. In this process, the material reacts with its surrounding environment, resulting in the degradation or deterioration of the material. Significant impacts of corrosion include increased maintenance costs, decreased equipment efficiency, and losses in production [1],[3].

To address the problem of corrosion, the use of corrosion inhibitors has become a major focus to reduce its adverse effects [4]. To tackle corrosion problems, much recent research has led to the use of corrosion inhibitors made from organic compounds for industrial metals [5]. These compounds are chosen because they are considered environmentally friendly, cost-effective, and effective in preventing corrosion [6]. These organic compounds can create a molecular layer that effectively protects metal surfaces from direct attack by corrosive substances, which is one of the main reasons for their use [7].

Pyridazine-derived compounds attract attention because they have diverse molecular structures, including functional groups such as sulfur, nitrogen, and oxygen that affect the adsorption ability of inhibitors on steel surfaces [3]. These compounds have been widely studied in the context of corrosion inhibition through various experimental methods such as gravimetry, potentiodynamic, and impedance spectroscopy [8].

Although experimental research provides valuable insights, it often requires significant time, cost, and resources [9]. As an alternative, Machine Learning (ML) approaches are becoming increasingly popular for evaluating corrosion inhibitor compounds [10]. ML facilitates the development of predictive models based on quantitative relationships between molecular structures and their properties and activities [11], [12]. Predictive models that have been developed allow prediction of the level of corrosion inhibition effectiveness of pyridazine-derived compounds [3].

Multilinear Regression (MLR) is a popular method in regression analysis that is effective for describing the linear relationship between input (X) and output (Y)

variables [13]. MLR works by determining the best line that represents the relationship between variables X and Y [14]. The main advantage of MLR is its ease of interpretation, which helps us understand the significant influence of input variables on output variables [3].

Adaptive Boosting Regression (ABR) often referred to as AdaBoost, is an ensemble learning method that improves model performance by combining several weak models into one strong model [1], [15]. ABR works by giving weight to each model used, where models that have low performance will be given a higher weight to be improved in the next iteration [16]. The advantage of ABR is its ability to handle overfitting and improve prediction accuracy with different model combinations [17], [15].

Many studies on the effectiveness of pyridazine compounds as corrosion inhibitors have utilized experimental methods, including gravimetry, potentiodynamics, and impedance spectroscopy [8], [18]. However, these experimental methods often require a lot of cost, time, and resources [19], [20]. As a solution, Machine Learning (ML) approaches offer a more efficient way of evaluating corrosion inhibitor performance. Using ML, it is possible to predict the effectiveness of inhibitors by identifying quantitative relationships between a compound's structure and its molecular properties [21], [14], [22]. Research by [14] shows that the artificial neural network (ANN) model has better performance than multilinear regression (MLR), with RMSE, MSE, and MAPE values of 111.5910, 10.5637, and 10.2362, respectively. Another study by Haikal et al. showed that the Decision Tree Regression (DTR) model is more effective than MLR, with R^2, MAE, and RMSE values of 0.908, 2.059, and 2.704, respectively [3].

However, our model achieved even lower error metrics in this study, making it more accurate and effective. Specifically, our model reached R^2, MSE, RMSE, and MAE values of 0.990, 0.004, 0.057, and 0.040, respectively, significantly outperforming the results from [14] and [3]. This improvement in model accuracy indicates the effectiveness of our approach in predicting the corrosion inhibition efficiency of pyridazine-derived compounds, highlighting the potential of more refined ML models in this field [23]. Therefore, this study aims to assess and compare the performance of Multilinear Regression (MLR) and AdaBoost Regression (ABR) models in forecasting the corrosion inhibition efficiency of pyridazine-derived compounds. This study implemented data normalization techniques in the pre-processing phase and used k-fold cross-validation using k=10 in the Machine Learning (ML) model-building process [24], [13]. the results of this study can provide a more accurate model and can be an important contribution to the development of more effective corrosion inhibitors in the future [6].

II. METHOD

Fig 1 illustrates the process of developing this study's Machine Learning (ML) model. The initial stage was the selection of the pyridazine dataset, followed by data preprocessing to address scale differences and sensitivity to outliers [3]. After that, the ML algorithm was selected to model the relationship between input and output variables that can accurately predict the corrosion inhibition efficiency of the pyridazine dataset. Furthermore, the model was trained using the k-fold cross-validation technique, which helps avoid overfitting and obtain a more generalized model. Evaluation of model performance was done using several metrics that are often used in regression model evaluation, namely Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2). These steps not only helped select the optimal ML model but also ensured that the model could produce accurate predictions of the corrosion inhibition efficiency of pyridazine in this study.

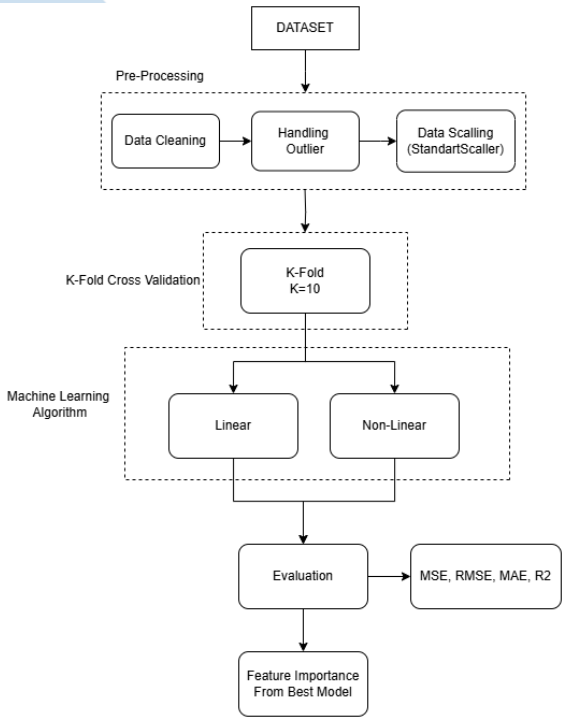

Fig 1. Development of Machine Learning Model

A. Dataset

The dataset used in this study is a collection of pyridazine-derived compound data published by Quadri et al. (2022) [14]. This data collection comprises 20 pyridazine compounds with various quantum molecular properties, used as independent variables and CIE values as dependent variables. The features included in this analysis are total energy (TE), HOMO (Highest Occupied Molecular Orbital), LUMO (Lowest Unoccupied Molecular Orbital), energy gap (ΔE), dipole moment (μ), ionization potential (I), electron

affinity (A), electronegativity (χ), global hardness (η), global softness (σ), and fraction of transferred electrons (ΔN).

B. Data Preprocessing

The initial stage of developing a Machine Learning (ML) model is to perform data normalization. This normalization process is fundamental in the pre-processing stage, as it ensures that all features are on a uniform scale. This helps to improve the performance of ML algorithms, especially if the features in the dataset have a wide range of values. In this study, the StandardScaler technique is used for normalization, transforming the data to have a mean of 0 and a standard deviation of 1. This approach effectively minimizes the impact of outliers, allowing the algorithm to perform more optimally [25], [13]. In addition, researchers applied the Interquartile Range (IQR) method to remove outlier data [18]. This step was taken to cleanse the dataset of unusual or unrepresentative data that could interfere with interpreting the analysis results. By calculating the IQR, researchers can identify and eliminate extreme data that may hurt model performance [6], [19].

C. K-Fold Cross-Validation

Researchers applied the k-fold cross-validation method with k = 10 to split the data into 10 equal parts [6]. At each iteration, the model was trained with 9 subsets and tested with the remaining subsets, alternating through all subsets [1]. This process aims to identify the model with the least error rate, thus ensuring model robustness and reliability [24]. The selection of k = 10 is done to maximize the use of data and minimize bias and variance [23]. The selection of k=10 was done to maximize the use of data and minimize bias and variance [23]. The selection of the k-fold value is tailored to the dataset's characteristics, although k=5 or k=10 values are commonly used in machine learning. This study evaluated the model's performance across various fold values from 1 to 10. After testing, it was determined that k=7 provided the best accuracy for this dataset. Therefore, we selected k=7 for the final model, reported in the results section. This explanation is included here to clarify the choice of k and its rationale. These steps were carefully designed to ensure the dataset's quality, with data preprocessing as a foundational step.

The ultimate goal is to create an accurate and consistent machine-learning model for predicting the corrosion inhibition efficiency of pyridazine derivatives [6]. Therefore, data preprocessing is an important basis in building a reliable and relevant model for this study.

D. Machine Learning Algorithm

In this research, various linear and non-linear regression algorithms have been applied to project CIE values [3]. This study involves a comparison between linear algorithms, which typically utilize linear relationships between input and output variables, and non-linear algorithms that can manage more complex and non-linear interactions between the two variable [6],[10]. Linear algorithms assessed include multilinear regression (MLR), ridge, lasso, Elastic-Net (EN), Support Vector Regression (SVR), and Generalized Linear Model (GLM), all prioritizing a linear relationship between input and output variables [6]. On the other hand, non-linear algorithms explored include random forest (RF), k-nearest neighbors (KNN), nu-support vector regressor (NuSVR), decision tree regressor (DT), gradient boosting regressor (GBR), orthogonal matching pursuit (OMP), kernel ridge (KR), partial least square (PLS), AdaBoost regressor (ABR), and bagging regressor (BR), all of which are designed to address more complex and non-linear relationships between input and output variables [1], [3], [6].

Through these experiments, the main objective was to identify the most effective algorithm to estimate how effective corrosion inhibition by pyridazine-derived compounds is [26]. The results show that the best model for predicting the efficiency of corrosion inhibitors is the AdaBoost Regressor (ABR). ABR is a boosting algorithm used to improve the accuracy of predictive models.[10]. The prediction of the AdaBoost Regressor is based on a combination of several weak learners [27]. Each weak learner contributes to the final prediction through a weighted vote, where the weight reflects the accuracy of the weak learner [28].

Multilinear Regression is a statistical technique for modeling linear relationships between input (X) and output (Y) variables:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon$$

where β is the regression coefficient that shows the effect of each independent variable X on the dependent variable$Y$, and? is the error term [29]. The MLR model was implemented using the scikit-learn library in Python.

AdaBoost Regression is an ensemble learning technique that aims to improve prediction accuracy by combining several weak models into one strong model [28],[6]. ABR works by assigning weights to each observation, and models that have high prediction errors get a more considerable weight in the next iteration [30]. The ABR algorithm is implemented using the scikit-learn library in Python.

The final prediction of H(x) from the AdaBoost Regressor is calculated as follows:

$$H(x) = \sum_{m=1}^{M} \alpha_m h_m(x)$$

where H(x) is the final prediction of the ensemble model,h_m is the weight of the mth weak learner, h_m is the prediction of the mth weak learner for input x, and M is the total number of weak learners in the ensemble [6]. The weighta_m is calculated based on the accuracy of each weak learner.

E. Model Evaluation

Model performance is evaluated using three main metrics, namely the coefficient of determination (R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE), which are evaluation metrics used to assess model performance [12], [31], [32]. The coefficient of determination (R^2) measures the effectiveness of independent variables in explaining changes in dependent variables, with values between 0 and 1, where values closer to 1 indicate that the model is more accurate [27], [33]. The MAE gives the average absolute error between the predicted and true values, providing a more intuitive understanding of the error rate [34]. RMSE measures the average prediction error but provides a larger penalty for more significant errors [10]. RMSE values closer to 0 indicate a more accurate model [30], [35]. By implementing k-fold cross-validation and utilizing these evaluation metrics, this study aims to provide a deep insight into the effectiveness of Multilinear Regression and AdaBoost Regression models in predicting the corrosion inhibition efficiency of pyridazine-derived compounds [3], [19].

F. Important Features

The analysis of important features in this study is a crucial step that requires an in-depth understanding of the relationship between the molecular and physicochemical characteristics of pyridazine-derived compounds and their corrosion inhibition efficiency [33]This method not only helps identify the most influential features in the prediction but also provides deeper insights into the corrosion-inhibiting mechanism of action of such compounds [33].[3]. Thus, the important feature analysis results provide a solid basis for selecting the most relevant features to construct accurate and effective Machine Learning models [27]. The information obtained from salient feature analysis also has important implications in the context of industrial applications and further research [6]. The discovery of the most significant features in demonstrating corrosion inhibition effectiveness can be used to direct the development of more cost-effective and environmentally friendly corrosion-inhibiting materials [36]. In addition, a deeper understanding of the relationship between the molecular properties of pyridazine-derived compounds and their performance as corrosion inhibitors provides opportunities for further research in the optimization of inhibitor formulations that can be widely applied in the corrosion industry [37]. Therefore, the analysis of important features is not only an important part of this research but also results in important impacts on the development of knowledge and technology in the effort to control corrosion [3], [6], [33].

III. Results and Discussions

The results of this study, including a comparison of the performance of Multilinear Regression (MLR) and AdaBoost Regression (ABR) models in estimating the corrosion inhibition effectiveness (CIE) of pyridazine-derived compounds, are illustrated in Table 1. This table presents the results of both models based on evaluation metrics such as the coefficient of determination (R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

TABLE I. Model Performance

Model	MLR and ABR model evaluation			
	MSE	***RMSE***	***MAE***	***R_2***
MLR	0.006	0.078	0.063	0.982
ABR	0.004	0.057	0.040	0.990

The ABR model shows an R^2 value closer to 1 and smaller MAE and RMSE values compared to the MLR model, proving that ABR has better prediction performance.

Figures 2, 3, 4, and 5 present the prediction distribution visualization and residual error plots for each model. These two types of visualizations are used to conduct a more detailed analysis of each model's accuracy.

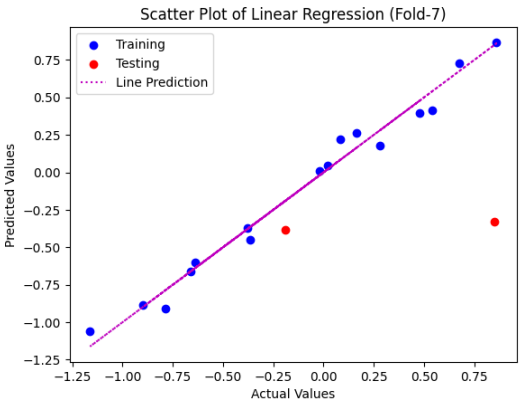

Fig 2. Scatter plot of MLR

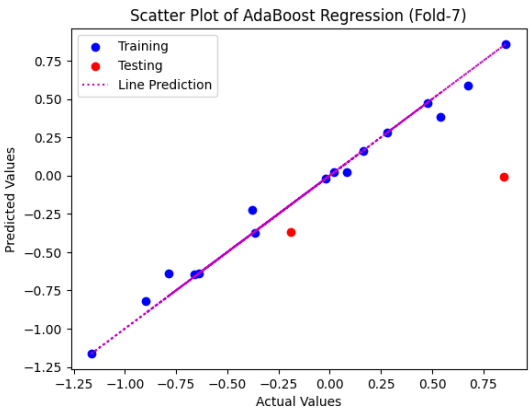

Fig 3. Scatter plot of ABR

In Figures 2 and 3, from the observed results, the distribution of prediction data for the AdaBoost Regressor (ABR) model looks closer to the line representing the true value (fitting line) compared to Multilinear Regression (MLR). This indicates that the ABR model has better prediction performance in processing the pyridazine derivative compound dataset. This means that the ABR model can produce more accurate predictions and closer to the original value than MLR. In addition, Figures 4 and 5 display the residual error, which is the difference between the actual and predicted values, showing that the distribution of prediction errors for the ABR model is closer to the 0 line than the MLR model. This shows that ABR produces smaller forecast errors, indicating higher accuracy in this model.

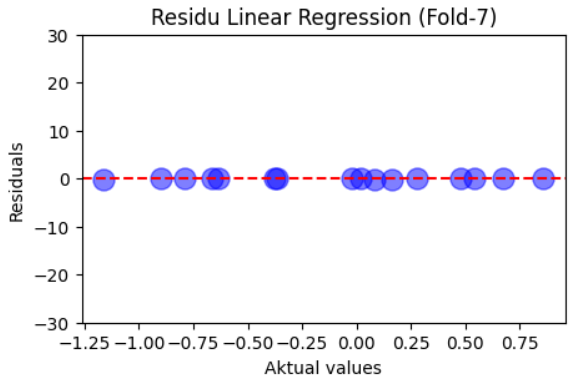

Fig 4. Residual Error on MLR model

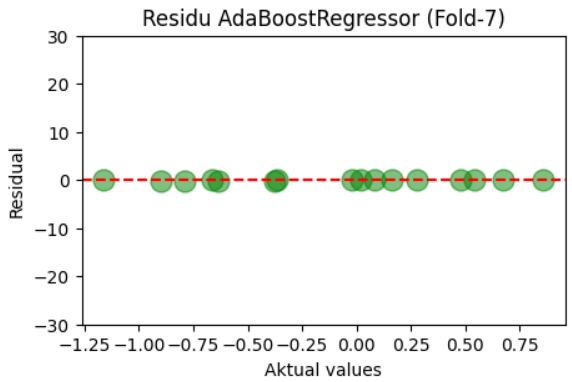

Fig 5. Residual Error on ABR model

Fig 6 shows the performance comparison between Multilinear Regression (MLR) and AdaBoost Regressor (ABR) based on R2 values. From the results shown, it can be seen that the ABR model consistently achieves R2 values close to 0.99 across all folds evaluated, demonstrating its ability to consistently explain around 95% of the variability in the data. In contrast, the MLR model showed more significant variation in R2 values, ranging from 0.82 to 0.98. This analysis suggests that the ABR model has a higher and more consistent accuracy in predicting pyridazine efficiency than the MLR model. The ability of the ABR model to consistently achieve R2 values close to 1 indicates that it can represent the relationship between input and output variables with a high degree of accuracy. Therefore, based on this evaluation, the ABR model is considered a more suitable choice to predict pyridazine derivatives' corrosion inhibition efficiency. This is supported by the ABR model's ability to provide accurate and consistent predictions and explain the variations in the evaluation data.

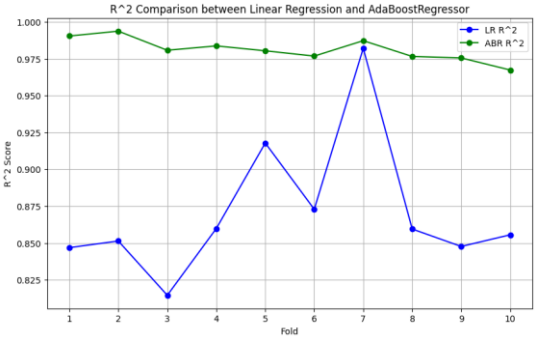

Fig 6. Comparison of the best linear (MLR) and nonlinear (ABR) algorithms for 10-fold trials.

The selection of (ABR) as the superior model is reinforced by the results of the feature importance analysis, which identifies key features that influence the prediction of (CIE). The important features found were total energy (TE), LUMO, dipole moment (μ), global softness (σ), HOMO, and ionization potential (I). Further analysis in Figure 12 shows that total energy (TE) and LUMO are the two most significant features in the (ABR) model. This shows that (ABR), with its ability to capture complex patterns, is better than other models such as (MLR) in predicting (CIE), making it a more appropriate choice for corrosion inhibition prediction.

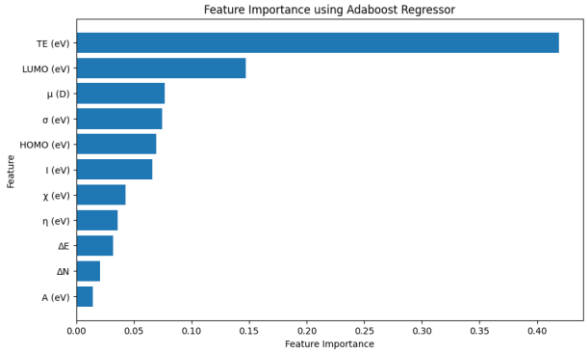

Fig 7. Feature Importance of ABR

IV. Conclusion

This study compares multilinear regression (MLR) and AdaBoost regression (ABR) algorithms to predict the corrosion inhibition effectiveness (CIE) of pyridazine derivatives. The results show that ABR is superior to MLR, with higher R^2 values (0.990 vs. 0.982) and smaller MAE and RMSE values. ABR also showed superior predictive performance and more effectively captured the complex relationship between molecular features and CIEs.

Feature analysis identified total energy (TE) and dipole moment (μ) as significant factors influencing CIE. The research methodology, including k-fold cross-validation and evaluation metrics, ensured robust and reliable results. These findings demonstrate the potential of ensemble models such as ABR in improving the prediction and efficiency of corrosion inhibitors, making an important contribution to developing more effective and environmentally friendly inhibitors.

ACKNOWLEDGMENT

The researchers would like to thank Allah SWT for His abundance of grace, making it easy to complete this research. We would also like to thank the Supervisor, Research Center for Quantum Computing and Materials Informatics, Faculty of Computer Science, Dian Nuswantoro University, for the support and facilities provided during this research process. We also want to express our deepest gratitude to our parents for their unending prayers, support, and motivation during this research process. Hopefully, the results of this research will be useful for the advancement of science and the welfare of mankind

REFERENCES

[1] Adryan Putra Sumarjono, M. Akrom, and G. Alfa Trisnapradika, "Comparison of the Best Machine Learning Model to Predict Corrosion Inhibition Capability of Benzimidazole Compounds," 2023.

[2] Ramadhan and T. L. Pratiwi, "SUPPRESSION OF CORROSION RATE OF ZINK IN HCl SOLUTION Utilization of Plants as Organic Inhibitors for Devotion Of Corrosion Rate Of Zinc in Hydrochloric Acid Utilization of Plants As Organic Inhibitors ... (Aditia Ramadhan, et al.)," vol. 14, no. 1, pp. 1-8.

[3] M. Haikal, M. Akrom, and G. A. Trisnapradika, "Comparison of Multilinear Regression and Decision Tree Regressor Algorithms in Predicting Pyridazine Corrosion Inhibition Efficiency," *Edumatic J. Educ. Inform.*, vol. 7, no. 2, pp. 307-315, Dec. 2023, doi: 10.29408/edumatic.v7i2.22127.

[4] Budi, M. Akrom, G. A. Trisnapradika, and T. Sutojo, "Optimization of Polynomial Functions on the NuSVR Algorithm Based on Machine Learning: Case Studies on Regression Datasets," vol. 10, no. 2, pp. 151-158, 2023, doi: 10.15294/sji.v10i2.43929.

[5] G. Against and L. Corrosion, "Effect of water hyacinth extract organic inhibitor addition on corrosion rate," vol. 2, no. 2, 2018.

[6] Y Mulyana, M. Akrom, and G. A. Trisnapradika, "Comparison of Linear and Non-Linear Machine Learning Algorithms for Predicting the Effectiveness of Plant Extracts as Corrosion Inhibitors," vol. 11, no. 1, 2024.

[7] M. Akrom, DFT Investigation of Syzygium Aromaticum and Nicotiana Tabacum Extracts as Corrosion Inhibitor, Science Tech: Jurnal Ilmu Pengetahuan dan Teknologi, 8(1), 42-48 (2022), doi: https://doi.org/10.30738/st.vol8.no1.a11775.

[8] E. H. R. Salim, M. M. B. Hammouti, and D. S. C. S. M. Almutairi, "Electrochemical Studies on New Pyridazinium Derivatives as Corrosion Inhibitors of Carbon Steel in Acidic Medium,"*J. Bio- Tribo-Corrosion*, vol. 0, no. 0, p. 0, 2019, doi: 10.1007/s40735-018-0195-3.

[9] M.Akrom, T. Sutojo, Investigasi Model Machine Learning Berbasis QSPR pada Inhibitor Korosi Pirimidin Investigation of QSPR-Based Machine Learning Models in Pyrimidine Corrosion Inhibitors, Eksergi, 20(1), 2023.

[10] V. Putranto, M. Akrom, and G. A. Trinapradika, "Implementation of Polynomial Function in Gradient Boosting Regressor Algorithm: Regression Study on Expired Medicines Dataset as Anticorrosion Material,"*J. Technol. and Manaj. Inform.*, vol. 9, no. 2, pp. 172-182, 2023, doi: 10.26905/jtmi.v9i2.11192.

[11] S. Budi, M. Akrom, H. Al Azies, U. Sudibyo, T. Sutojo, G. A. Trisnapradika, A. N. Safitri, A. Pertiwi, and S. Rustad, Implementation of polynomial functions to improve the accuracy of machine learning models in predicting the corrosion inhibition efficiency of pyridine-quinoline compounds as corrosion inhibitors, KnE Eng., pp. 78-87, 2024.

[12] Leni, "Selection of Optimal Machine Learning Algorithm for Prediction of Mechanical Properties of Aluminum,"*J. Energy Engines, Manufacturing, and Mater.*, vol. 7, no. 1, p. 35, 2023, doi: 10.30588/jeemm.v7i1.1490.

[13] Ambarwari, Q. Jafar Adrian, and Y. Herdiyeni, "Analysis of the Effect of Data Scaling on the Performance of the Machine Learning Algorithm for Plant Identification,"*J. RESTI (Engineering Systems and Technol. Information)*, vol. 4, no. 1, pp. 117-122, 2020, doi: 10.29207/resti.v4i1.1517.

[14] W. Quadri*et al.*, "Development of QSAR-based (MLR/ANN) predictive models for effective design of pyridazine corrosion inhibitors," *Mater. Today Commun.*, vol. 30, no. October 2021, p. 103163, 2022, doi: 10.1016/j.mtcomm.2022.103163.

[15] Juni, "Comparison of Linear and Non- Linear Machine Learning Algortima for Predicting the Effectiveness of Plant Extracts as Corrosion Inhibitors," 2024.

[16] Irma Prianti, "Comparison of K-Nearest Neighbor and Adaptive Boosting Methods in Multi-class Classification Cases,"*J Stat. J. Ilm. Theor. and Appl. Stat.*, vol. 13, no. 1, pp. 39-47, 2020, doi: 10.36456/jstat.vol13.no1.a3269.

[17] B. Aji, "Faculty of Engineering - Muria Kudus University153," *Pros. 4th SNATIF 2017*, no. 2015, pp. 153-160, 2017.

[18] Luo*et al.*, "A new pyridazine derivative synthesized as an efficient corrosion inhibitor for copper in sulfuric acid medium: Experimental and theoretical calculation studies," *J. Mol. Liq.*, vol. 341, p. 117370, 2021, doi: 10.1016/j.molliq.2021.117370.

[19] Akrom and T. Sutojo, "Investigation of QSPR-Based Machine Learning Models in Pyrimidine Corrosion Inhibitors," vol. 20, no. 2, pp. 107-111, 2023.

[20] T. Ser, Ž. Petar, and M. W. Wong, "Applied Surface Science Prediction of corrosion inhibition e ffi ciency of pyridines and quinolines on an iron surface using machine learning-powered quantitativestructure- property relationships," vol. 512, no. February, 2020, doi: 10.1016/j.apsusc.2020.145612.

[21] Hassan*et al.*, "Heliyon Development and validation of QSPR models for corrosion inhibition of carbon steel by some pyridazine derivatives in acidic medium," vol. 6, no. April, 2020, doi: 10.1016/j.heliyon.2020.e05067.

[22] Sutojo, "OPEN A machine learning approach for corrosion small datasets," 2023, doi: 10.1038/s41529-023-00336-7.

[23] antoni wibowo, "10 Fold-Cross Validation," vol.

(accessed, 2003.

[24] Khoirunnisa and N. G. Ramadhan, "Improving malaria prediction with ensemble learning and robust scaler: An integrated approach for enhanced accuracy," *J. Infotel*, vol. 15, no. 4, pp. 326-334, 2023, doi: 10.20895/infotel.v15i4.1056.

[25] Subhi*et al.*, "Implementation of Scaling Techniques in a Website-Based Server Balancing Management System," *J. Comput. and Appl.*, vol. 09, no. 02, pp. 316-326, 2021.

[26] R. Rosyid, L. Mawaddah, and M. Akrom, "Investigation of Regression Machine Learning Models on Medicinal Compounds as Corrosion Inhibitors," no. Ml, 2024, doi: 10.33364/algoritma/v.21-1.1598.

[27] Anjum, K. Khan, W. Ahmad, A. Ahmad, M. N. Amin, and A. Nafees, "Application of Ensemble Machine Learning Methods to Estimate the Compressive Strength of Fiber-Reinforced Nano-Silica Modified Concrete,"*Polymers (Basel).*, vol. 14, no. 18, 2022, doi: 10.3390/polym14183906.

[28] P. Solomatine and D. L. Shrestha, "AdaBoost.RT: A boosting algorithm for regression problems,"*IEEE Int. Conf. Neural Networks - Conf. Proc.*, vol. 2, no. August 2004, pp. 1163-1168, 2004, doi: 10.1109/ijcnn.2004.1380102.

[29] Wulandari, A. Perdana Windarto, H. Qurniawan, and G. Artikel, "Application of Multiple Regression in Estimating the Amount of Population Growth in Siantar District Article Info ABSTRACT,"*JOMLAI J. Mach. Learn. Artif. Intell.*, vol. 1, no. 4, pp. 2828-9099, 2022, doi: 10.55123/jomlai.v1i4.1677.

[30] A. Mortara, M. Permatasari, A. Desiani, Y. Andriani, and M. Arhami, "Comparison of C4.5 and Adaptive Boosting Algorithms in Alzheimer's Disease Classification," *J. Technol. and Inf.*, vol. 13, no. 2, pp. 196-207, 2023, doi: 10.34010/jati.v13i2.10525.

[31] Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation,"*PeerJ. Comput. Sci.*, vol. 7, p. e623, 2021, doi: 10.7717/peerj-cs.623.

[32] O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," no. 2, pp. 5481-5487, 2022.

[33] Leng*et al.*, "Machine learning predicting and engineering the yield, N content, and specific surface area of biochar derived from pyrolysis of biomass," *Biochar*, vol. 4, no. 1, 2022, doi: 10.1007/s42773-022-00183-w.

[34] A. Suryanto, "Application of Mean Absolute Error (Mea) Method in Linear Regression Algorithm for Rice Production Prediction,"*Saintekbu*, vol. 11, no. 1, pp. 78-83, 2019, doi: 10.32764/saintekbu.v11i1.298.

[35] Dzisofi Amelia, S. Wahyuni, and D. Harisuseno, "Evaluation of Satellite Data Suitability as an Alternative to Evaporation Data Availability in Wonorejo Reservoir,"*J. Tech. Irrigation*, vol. 12, no. 2, pp. 127-138, 2021, doi: 10.21776/ub.pengairan.2021.012.02.05.

[36] Akrom and C. Science, "Investigation Of Natural Extracts As Green Corrosion Inhibitors In Steel Using Density Functional Theory Extracts of Natural Products as Green Corrosion Inhibitors on Steel Nowadays, the development of," vol. 10, no. 01, pp. 89-102, 2022.

[37] Nayem, A. C. Mohammad, and K. Mohamed, "An overview of green corrosion inhibitors for sustainable and environment friendly industrial development An overview of green corrosion inhibitors for sustainable and environment ... An overview of green corrosion inhibitors for sustainable and environment friendly industrial Nayem Hossain, Mohammad Asaduzzaman Chowdhury & Mohamed,"*J. Adhes. Sci. Technol.*, vol. 0, no. 0, pp. 1-18, 2020, doi: 10.1080/01694243.2020.1816793