Jurnal Ilmu Komputer dan Informasi (Journal of Computer Science and Information)
18/1 (2025), 1-9. DOI: http://dx.doi.org/10.21609/jiki.v18i1.1303

Transformative Insights into Corrosion Inhibition: A Machine Learning Journey from Prediction to Web-Based Application

Dzaki Asari Surya Putra[1], Nicholaus Verdhy Putranto[1], Nibras Bahy Ardyansyah[1], Gustina Alfa Trisnapradika[1,2], Muhamad Akrom[1,2*]

[1]Study Program in Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, 50131, Indonesia
[2]Research Center for Materials Informatics, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, 50131, Indonesia

Correspondence: m.akrom@dsn.dinus.ac.id

Abstract

This study focuses on the exploration and evaluation of machine learning (ML) models to analyze expired pharmaceutical data for their potential use as corrosion inhibitors. Additionally, the entire modeling process is integrated into a user-friendly platform through a Streamlit service-assisted corrosion inhibitor website, facilitating broader accessibility and practical application. The models are trained offline to ensure accurate performance, eliminating the need for users to retrain the models themselves. This approach simplifies the user experience by offering a ready-to-use prediction service directly on the website platform. Among the various ML models implemented, XGB demonstrated the highest performance with an R2-score of 0.99999999. Given that many chemists are not familiar with informatics coding, the researchers developed a Streamlit-based website that includes tools to customize the models. The end product of this work is a corrosion inhibitor experimentation tool that eliminates the need for users to code, making advanced ML techniques accessible to a broader audience within the chemistry community.

Keywords: *corrosion, inhibitor, machine learning, web, streamlit*

1. Introduction

Corrosion represents the degradation of metal, characterized by chemical or electrochemical processes occurring between a metal and its surrounding environment, leading to a deterioration in metal quality [1], [2], [3]. Its significance in the environmental context is paramount, and corrosion manifestations can vary depending on the environmental conditions to which a metal is exposed, making consequences challenging to discern. The occurrence of corrosion is influenced by factors such as metal reactivity, degradation, air presence, humidity levels, gas coefficients (e.g., sulfur dioxide and carbon dioxide), and the existence of electrolytes [3]. Corrosion results in losses across diverse sectors, impacting industrial assets, buildings, and various infrastructures [4]. Consequently, the utilization of corrosion inhibitors is imperative to address and mitigate existing corrosion challenges effectively [5], [6].

Corrosion inhibitors play a pivotal practical role, frequently employed to diminish metal losses in the production process and mitigate the potential for material failure [7], [8]. The occurrence of these two events can result in abrupt disruptions in industrial operations, ultimately incurring additional costs. Research indicates that inhibitors, including drugs, play a crucial role in diverse environments and at various concentration levels, exhibiting effectiveness against a range of metals such as zinc, copper, mild steel, aluminum, bronze, and carbon steel. These inhibitors demonstrate the capability to create a protective layer on material surfaces, preventing corrosive reactions between the material and the environment and, consequently, mitigating or reducing corrosion levels in specific materials [9].

Conventional methods for determining the effectiveness of corrosion inhibitors are slow and costly due to the need for chemical experts, materials, and equipment. These methods typically involve experimental techniques such as weight loss assessment, potentiodynamic polarization, and electrochemical impedance spectroscopy [10]. The economic toll of addressing corrosion-related issues amounts to approximately USD 2.5 trillion annually. We propose using machine learning (ML) as a faster, more cost-effective alternative. ML models analyze historical data to predict inhibitor performance, reducing the need for extensive laboratory experiments and chemical

materials. This approach provides a quicker, economical method to predict Corrosion Inhibition Efficiency (CIE) through computation.

ML, a subset of artificial intelligence (AI), empowers systems to learn from data, discern both linear and non-linear relationships, and enhance performance without necessitating intricate programming by humans [11], [12]. In the context of predicting and analyzing CIE, the ML approach is employed to construct a predictive model capable of estimating the extent of corrosion in materials by considering influencing attributes [6]. Previous research demonstrates that models utilizing nonlinear algorithms outperform those employing linear algorithms in predicting the corrosion inhibition capability of Benzimidazole compounds, as evidenced by superior R^2, MAE, MSE, and RMSE values [13]. Other investigations have compared the performance of two ML algorithms in predicting the efficiency of pyridazine corrosion inhibitors. Results indicate that the Decision Tree regressor (DT) algorithm outperforms the multilinear regression algorithm, as evidenced by higher R^2, MAE, and RMSE values associated with the DT algorithm [14]. Additional research, employing nonlinear algorithms on datasets of expired corrosion inhibitor drugs using a Gradient Boosting Regressor (GBR) with a polynomial function, degrades the RMSE value of previous research from 3,89% to 0,03%. GBR is reflected in an R^2 value of 0.9998 for training data and 0.9999 for testing data, highlighting the model's proficiency in predicting CIE [6].

In this research, our objective is to develop an application that efficiently aids users in estimating the CIE as a corrosion inhibitor in expired drug compounds. Through the implementation of both linear and nonlinear algorithms on the corrosion inhibitor website, we anticipate that scientists can enhance the efficiency of predicting the CIE of compounds and provide valuable information regarding expired drugs in the context of corrosion inhibitors [15]. There are two primary features on this website: the Customizing Model feature allows users to experiment with different models and data treatments without having to perform information coding, and the best model feature provides information about the machine learning process so that you can obtain an R^2 of 0.9998 [6].

While our focus primarily revolves around predicting CIE outcomes through website development utilizing ML, there are related works that employ similar approaches albeit with differing focal points. One such example is a study by Ahmed et al [16] where they utilized ML algorithms within website development to predict diabetes. Despite not specifically targeting CIE prediction, their research underscores the efficacy of integrating ML techniques into website development for diverse objectives. Although the primary focus of Ahmed et al work differs from ours, their utilization of ML within website development shares common ground in leveraging data-driven methodologies to optimize website performance and user experience. This parallel underscores the versatility and applicability of ML techniques in diverse domains, including website development, beyond our specific focus on CIE prediction. Therefore, while distinct in objectives, both studies collectively contribute to the broader exploration of ML applications in web development, highlighting the multifaceted capabilities of ML algorithms in optimizing digital experiences.

2. Method

The implementation of ML typically involves several stages, including data collection, data preprocessing, modeling, evaluation, and implementation. The dataset of drug compounds is gathered based on predefined descriptors. Data preparation is then conducted to make the information suitable for further ML processing. Utilizing predetermined algorithms, the modeling process is executed to train the data. Following this, the model undergoes an evaluation to assess its performance and accuracy in target prediction, utilizing available descriptors. Finally, Streamlit is employed to develop an online application that encapsulates the model. The research methodology is depicted in Figure 1, with a detailed explanation in sections 2.1 to 2.5.

2.1. Data Pre-processing

The preprocessing phase involves data separation to identify missing values and outliers [6]. To enhance the robustness of the ML model, data scaling is performed, ensuring each feature shares a consistent scale [18], [19]. The MinMax scaler technique is employed in this study. This scaling method rescales the values of each feature within a dataset to a common range. It involves subtracting the minimum value of each feature from all values in that feature and then dividing by the range of values in the feature.

$$Xscaled = \frac{(X - Xmin)}{(Xmax - Xmin)} \quad (1)$$

Where X_{min} is the minimum value in the feature X, and X_{max} is the maximum value in the feature X. MinMax Scaler transforms the data to fall within a specified range, typically between 0 and 1.

In this study, for the data splitting phase, we employed two cross-validation techniques:

HoldOut Cross Validation (HOCV) and K-Fold Cross Validation (KFCV). The HOCV technique involves splitting the dataset into training and testing sets, ensuring that each data point is exclusively assigned to either the training or testing set. It undergoes an accuracy assessment by training the model on the training set and evaluating its performance on the separate testing set, providing a comprehensive evaluation of the model's predictive capabilities [20]. Data will be split into ratios 80:20, 70:30, and 60:40. KFCV: Geared towards avoiding overfitting and obtaining a robust estimate of how well a model generalizes, KFCV iterates k times. In each iteration, it designates one chunk of the data as the validation set and trains the model on the remaining chunks [21]. With KFCV, the data will be split using the variations k = 3, k = 5, and k = 10.

The utilization of scaler techniques, cross-validation, and hyperparameter tuning on each model is pivotal. A scaler ensures that the features have a uniform scale, thereby enhancing the convergence speed and stability of the model. For instance, disparate scales among features may impede machine learning algorithms in correctly adjusting their weights. The scaler effectively normalizes or retransforms the features to attain a uniform distribution. Moreover, the application of K-fold cross-validation facilitates the objective evaluation of model performance by partitioning the data into K-equal segments and validating each fold. This method aids in averting overfitting while providing an estimation of the actual model performance. Additionally, every machine learning algorithm encompasses a plethora of hyperparameters that necessitate determination before model training. The judicious selection of appropriate hyperparameters significantly impacts model performance. By meticulously attending to these factors during model development and evaluation, the resultant model manifests optimal performance, thereby instilling confidence in its capability to address pertinent challenges.

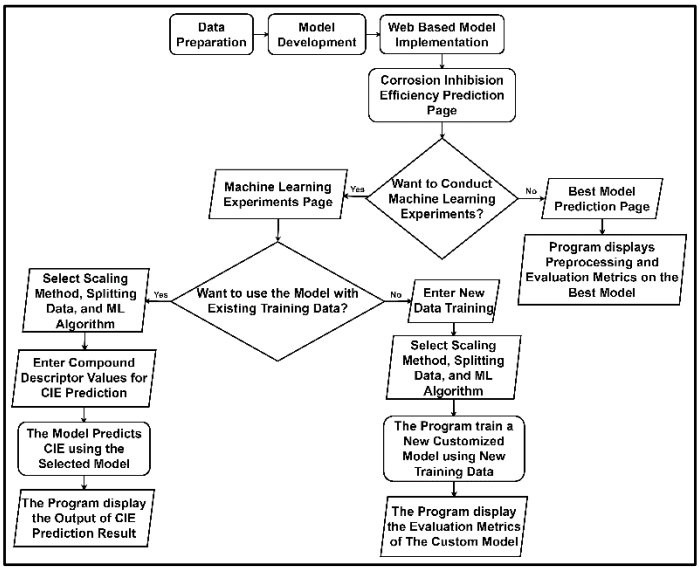

Figure 1. Research methodology

2.2. ML Modeling

The successful implementation of the Corrosion Inhibitor Website necessitates the development of an ML model capable of accurately predicting the CIE of a compound. This predictive capability enhances the selection process for scientists, aiding in the identification of the most effective corrosion inhibitors for specific applications. In the process of model construction, a diverse set of both linear and nonlinear algorithms is employed to accommodate the varying complexities present in the dataset [13].

In our research, we chose to implement both linear and nonlinear algorithms to comprehensively address the predictive task. The selected linear algorithms for implementation include Linear Regression (LR), Ridge, Lasso, and Elastic Net. These algorithms were chosen due to their ability to model the linear relationship between predictor (independent variable) and target (dependent variable). We ensured that these models underwent thorough scrutiny to validate their adherence to the basic assumptions associated with linear algorithms, as outlined in previous studies [22], [23]. Additionally, we opted to incorporate several nonlinear algorithms to capture more complex relationships between dependent (target) and independent (feature) variables. The nonlinear algorithms chosen for implementation encompass

AdaBoost, CatBoost, XGBoost (XGB), Light GBM (LGBM), Gradient Boosting Regressor (GBM), Random Forest (RF), Support Vector Regression (SVR), and Nu-SVR. These models were selected based on their demonstrated effectiveness in capturing nonlinear patterns and improving prediction accuracy through algorithmic iterations, as supported by prior research [15], [24].

By implementing a diverse range of algorithms, including both linear and nonlinear approaches, we aimed to explore the predictive task comprehensively and identify the most suitable model(s) for our specific dataset and objectives. This approach enables us to leverage the strengths of different algorithms and gain deeper insights into the underlying patterns within the data. Ultimately, the selection of these models was driven by the desire to maximize predictive accuracy and robustness while ensuring compatibility with the characteristics of our dataset and research objectives.

2.3. Model Evaluation

To enhance the model's effectiveness in delivering precise predictions on the Corrosion Inhibitor Website, we utilize regression metrics including R^2 (Coefficient of Determination) score and RMSE (Root Mean Squared Error). The selection of R^2 as a metric was primarily based on its widespread acceptance and interpretability, especially in regression tasks. R^2 provides a measure of the proportion of variance in the dependent variable that is predictable from the independent variables. When R^2 is close to 1, it indicates that a high percentage of the variance in the dependent variable is explained by the independent variables. The RMSE metric was chosen as a measure of model performance due to its widespread use in regression tasks, particularly in cases where the prediction errors are expected to be normally distributed. RMSE indicates the average magnitude of the residuals or prediction errors. Despite its mathematical simplicity, RMSE effectively penalizes larger errors more heavily than smaller ones, making it particularly suitable for applications where accurately estimating the magnitude of prediction errors is crucial. R^2 is employed as a measure of how well the predicted values align with the actual values, with a range of values from 0 to 1 [14]. When a model achieves a lower RMSE value, it means that the probability of prediction errors has been minimized [7]. The optimal ML model is characterized by an R^2 score approaching 1 and a minimum RMSE value [13], [25], [26]. This measure is used to evaluate the precision of the model since a lower degree of statistical error indicates a more accurate prediction model [7]. RMSE serves as a metric to quantify the accuracy of predictions by assessing the average squared difference between predicted values and actual values [14].

$$R^2 = \frac{\sum_{i=1}^{n}(Y_i' - \bar{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2} \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i' - \bar{Y}_i)^2} \quad (3)$$

Where Y_i represent the actual values, $\bar{Y}_i$ denotes the mean of actual values, and Y_i' indicates predicted values.

2.4. Implementation Using Streamlit

Upon completing the modeling phase, the subsequent step involves implementing the model in a web application using Streamlit, providing scientists with a convenient means to assess the CIE of compounds. Streamlit, an open-source Python-based web application framework, is chosen for its speed and flexible development capabilities. Typically employed for analysis, creation of intricate interactive applications, and implementation of ML models [10], [27], Streamlit serves as a valuable tool for designing ML user interfaces. In the design of the interface, researchers have created two options for predicting CIE. These alternatives encompass prediction based on a single input and prediction based on a CSV file containing descriptor features of drug compounds.

The website is also designed to handle incomplete datasets. When a user submits a dataset with missing values or incomplete variables, the preprocessing pipeline embedded within the website performs data imputation techniques to address missing values. This preprocessing step ensures that the dataset is complete and compatible with the trained ML models. Once the dataset is preprocessed and made complete, the user can then select the desired prediction task on the website interface. The website utilizes pre-trained ML models to make predictions based on the provided dataset. These predictions are generated in real time, leveraging the computational power of the server hosting the website. The predicted outcomes are presented to the user through the website interface, providing clear and interpretable results. Additionally, users have the option to further analyze or visualize the predictions using built-in tools or export the results for further processing. Periodically, the ML models hosted on the website may undergo updates or retraining using new data to ensure their relevance and accuracy. However, users are not required to participate in the model training process; instead, they benefit from the improved performance of the updated models seamlessly integrated into the website.

3. Result and Discussion

The RMSE values for each combination of linear and nonlinear algorithms were determined through data training and data splitting for each algorithm. The specific RMSE values for each combination are presented in Table 1 and Figure 2.

Table 1. RMSE and R^2 scores of linear and nonlinear algorithm combination in data testing

Algorithm	RMSE	R^2 Scores
XGB Regressor	0.00059064	0.99999999
GBR	0.00152432	0.99999994
CatBoost Regressor	0.01569679	0.99999308

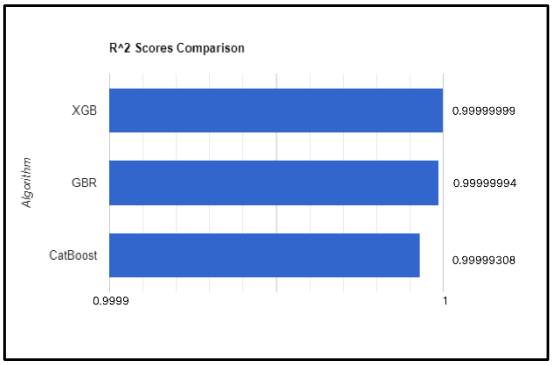

Figure 2. R^2 Scores of top 3 algorithm combinations in data testing

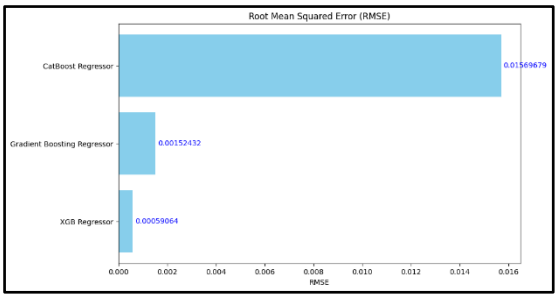

Figure 3. RMSE scores of top 3 algorithm combinations in data testing

It is clear from Table 1 and Figures 2 and 3 that the XGB Regressor yields the best results. Additionally, a visual representation of the prediction plot is presented, offering insights into the model's performance in forecasting CIE. The data visualization that follows demonstrates this. As illustrated in Figure 4.

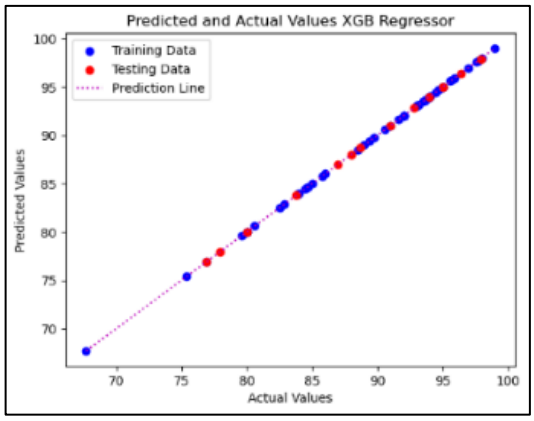

Figure 4. Data visualization for XGB Regressor

Based on the conducted testing, models displaying favorable R^2 and RMSE scores are predominantly associated with non-linear algorithm models. Therefore, these models are more recommended for predicting the CIE in the case study of inhibitor compounds [6], [13]. The non-optimal RMSE and R^2 scores in Table 1 indicate that the linear algorithm model tends to forecast CIE with an inadequate level of accuracy. Non-linear algorithms, on the other hand, may estimate a drug compound's potential CIE more accurately.

Once the ML modeling is complete, the entire modeling phase is implemented in a web-based application called Streamlit. This is illustrated in Figures 5 to 12.

Based on Figures 5 and 6, the machine learning modeling process of the best model to predict the CIE of a compound is presented. This page is a guide to building good machine learning to predict the CIE of a compound. It provides a platform for scientists to explore the application of ML in predicting the CIE of corrosion-inhibiting drug compounds. In Figure 7, options for ML experiments that can be conducted.

Figure 5. Model description page of Best Model.

Figure 6. Model description page of Best Model

Figure 7. Option to predict using existing models.

From Figure 8, the Model Customization page in the Machine Learning Experiment offers a variety of linear and non-linear algorithms that can be selected for prediction. This provides a platform for scientists to explore the application of ML in predicting the CIE of corrosion-inhibiting drug compounds. In addition, the option to select model customization based on algorithm, scaling, and data splitting.

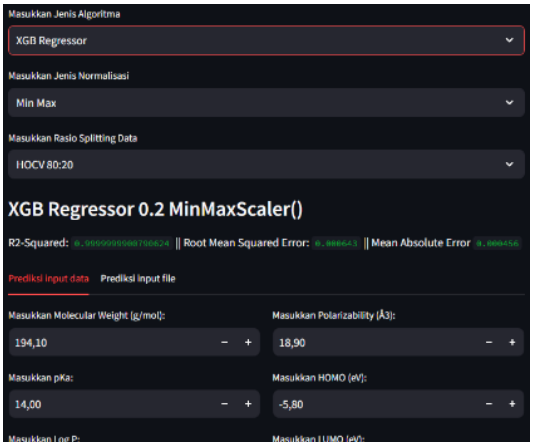

Figure 8. Selecting model combination based on algorithm, scaling, and data splitting.

In Figure 9, using this customization model option, scientists can easily achieve a combined CIE. These models are pre-trained using algorithms, data splitting, and scaling to improve the sensitivity analysis of ML models and standardize all features to the same scale [28], [29]. Scientists can then utilize the ML capabilities to predict a compound's CIE based on its features.

Figure 9. CIE prediction using a combination of customization models.

To ascertain the CIE of a compound, scientists only need to input data that characterizes the compound they want to predict. Apart from using existing models to predict CIE, scientists can also experiment to create Machine Learning models using new training data. Refer to Figure 10 for details.

Figure 10. Option to train training data using ML algorithms.

The ML Experiment option enables scientists to upload training drug compounds containing 10 molecular descriptor features. Subsequently, scientists can select the algorithm, scaling method, and data splitting approach to be applied in training the data, as depicted in Figure 11.

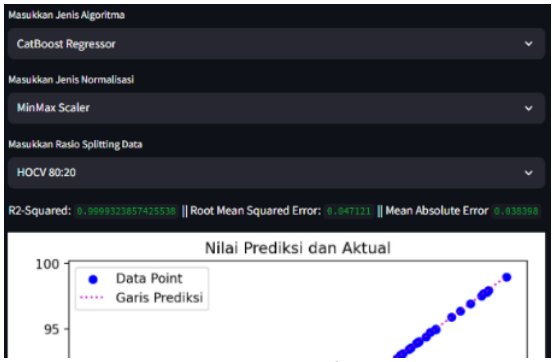

Figure 11. Model evaluation in R^2 Scores and RMSE.

After customizing the algorithm, scaling the features, and partitioning the dataset, the performance of the model will be evaluated using metrics such as R2 Score and RMSE. These metrics provide insight into how well the model captures the underlying pattern of the data and how accurately it predicts the target variable. In addition, a visualization of the prediction plot will be displayed as shown in Figure 12.

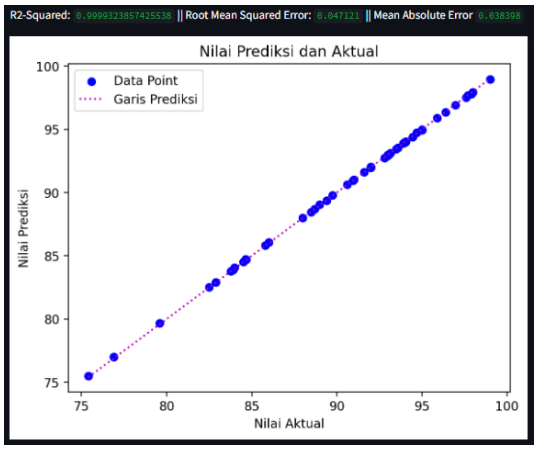

Figure 12. Visualization of the prediction plot of the training model input data.

4. Conclusion

The Corrosion Inhibitor Website offers a streamlined and efficient solution for predicting the CIE of expired pharmaceutical compounds. The successful application of ML within the Streamlit web-based platform demonstrates its effectiveness in predicting potential CIE. This implementation provides valuable support for researchers in efficiently calculating the CIE of drug compounds, thereby enhancing capacity in corrosion inhibition. Considering alternative algorithms or methods is crucial when implementing an ML approach for predicting the CIE of drug compounds, aiming to improve accuracy and prediction efficiency.

References

[1] M. Akrom, A.G. Saputro, A.L. Maulana, A. Ramelan, A. Nuruddin, S. Rustad, and H.K. Dipojono, DFT and microkinetic investigation of oxygen reduction reaction on corrosion inhibition mechanism of iron surface by Syzygium Aromaticum extract, *Appl Surf Sci*, 615, 156319 (2023), https://doi.org/10.1016/j.apsusc.2022.156319.

[2] S. Budi, M. Akrom, G.A. Trisnapradika, T. Sutojo, W.A.E. Prabowo, Optimization of Polynomial Functions on the NuSVR Algorithm Based on Machine Learning: Case Studies on Regression Datasets, *Scientific Journal of Informatics*, 10(2), (2023), https://doi.org/10.15294/sji.v10i2.43929.

[3] S. Tanwer and S. K. Shukla, "Recent advances in the applicability of drugs as corrosion inhibitor on metal surface: A review," *Curr. Res. Green Sustain. Chem.*, vol. 5, p. 100227, 2022, doi: 10.1016/j.crgsc.2021.100227.

[4] A. R. Permanasari, T. R. Saputra, A. N. Aina, and S. Liska, "Penentuan Pelarut Terbaik pada Ekstraksi Tanin Kulit Kayu Akasia dan Pengaruhnya Sebagai Inhibitor Laju Korosi pada Baja Karbon," *J. Tek. Kim. Dan Lingkung.*, vol. 4, no. 1, pp. 7–16, Apr. 2020, doi: 10.33795/jtkl.v4i1.129.

[5] M. Akrom *et al.*, "Artificial Intelligence Berbasis QSPR Dalam Kajian Inhibitor Korosi," *JoMMiT J. Multi Media Dan IT*, vol. 7, no. 1, pp. 015–020, Jul. 2023, doi: 10.46961/jommit.v7i1.721.

[6] N. V. Putranto, M. Akrom, and G. A. Trinapradika, "Implementasi Fungsi Polinomial pada Algoritma Gradient Boosting Regressor: Studi Regresi pada Dataset Obat-Obatan Kadaluarsa sebagai Material Antikorosi," 2023.

[7] M. Akrom, S. Rustad, and H. Kresno Dipojono, "Machine learning investigation to predict corrosion inhibition capacity of new amino acid compounds as corrosion inhibitors," *Results Chem.*, vol. 6, p. 101126, Dec. 2023, doi: 10.1016/j.rechem.2023.101126.

[8] M. Akrom, Investigation of natural extracts as green corrosion inhibitors in steel using density functional theory, *Jurnal Teori dan Aplikasi Fisika*, 10(1), 89-102 (2022), https://doi.org/10.23960%2Fjtaf.v10i1.2927.

[9] C. T. Ser, P. Žuvela, and M. W. Wong, "Prediction of corrosion inhibition efficiency of pyridines and quinolines on an iron surface using machine learning-powered quantitative structure-property relationships," *Appl. Surf. Sci.*, vol. 512, p. 145612, May 2020, doi: 10.1016/j.apsusc.2020.145612.

[10] M. Akrom, Green Corrosion Inhibitors for Iron Alloys: A Comprehensive Review of Integrating Data-Driven Forecasting, Density

Functional Theory Simulations, and Experimental Investigation, *Journal of Multiscale Materials Informatics,* vol. 1, pg. 22-37 (2024). https://doi.org/10.62411/jimat.v1i1.10495.

[11] M. Akrom, S. Rustad, H.K. Dipojono, Development of Quantum Machine Learning to Evaluate the Corrosion Inhibition Capability of Pyrimidine Compounds, *Mater Today Commun*, 39, 108758 (2024), https://doi.org/10.1016/J.MTCOMM.2024.108758.

[12] D. Leni, H. Yermadona, A. Usra Berli, R. Sumiati, and H. Haris, "Pemodelan Machine Learning untuk Memprediksi Tensile Strength Aluminium Menggunakan Algoritma Artificial Neural Network (ANN)," *J. Surya Tek.*, vol. 10, no. 1, pp. 625–632, Jun. 2023, doi: 10.37859/jst.v10i1.4843.

[13] C. A. P. Sumarjono, M. Akrom, and G. A. Trisnapradika, "Perbandingan Model Machine Learning Terbaik untuk Memprediksi Kemampuan Penghambatan Korosi oleh Senyawa Benzimidazole," *Techno.Com*, vol. 22, no. 4, pp. 973–980, Nov. 2023, doi: 10.33633/tc.v22i4.9201.

[14] F. M. Haikal, M. Akrom, and G. A. Trisnapradika, "Perbandingan Algoritma Multilinear Regression dan Decision Tree Regressor dalam Memprediksi Efisiensi Penghambatan Korosi Piridazin," *Edumatic J. Pendidik. Inform.*, vol. 7, no. 2, pp. 307–315, Dec. 2023, doi: 10.29408/edumatic.v7i2.22127.

[15] T. Sutojo, S. Rustad, M. Akrom, A. Syukur, G. F. Shidik, and H. K. Dipojono, "A machine learning approach for corrosion small datasets," *Npj Mater. Degrad.*, vol. 7, no. 1, p. 18, Mar. 2023, doi: 10.1038/s41529-023-00336-7.

[16] N. Ahmed *et al.*, "Machine learning based diabetes prediction and development of smart web application," *Int. J. Cogn. Comput. Eng.*, vol. 2, pp. 229–241, Jun. 2021, doi: 10.1016/j.ijcce.2021.12.001.

[17] C. Beltran-Perez *et al.*, "A General Use QSAR-ARX Model to Predict the Corrosion Inhibition Efficiency of Drugs in Terms of Quantum Mechanical Descriptors and Experimental Comparison for Lidocaine," *Int. J. Mol. Sci.*, vol. 23, no. 9, p. 5086, May 2022, doi: 10.3390/ijms23095086.

[18] Pelsri Ramadar Noor Saputra and A. Chusyairi, "Perbandingan Metode Clustering dalam Pengelompokan Data Puskesmas pada Cakupan Imunisasi Dasar Lengkap," *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 4, no. 6, Dec. 2020, doi: 10.29207/resti.v4i6.2556.

[19] M. Akrom, S. Rustad, A. G. Saputro, and H. K. Dipojono, "Data-driven investigation to model the corrosion inhibition efficiency of Pyrimidine-Pyrazole hybrid corrosion inhibitors," *Comput Theor Chem*, vol. 1229, p. 114307, Nov. 2023, https://doi.org/10.1016/J.COMPTC.2023.114307.

[20] F. Tempola, R. Rosihan, and R. Adawiyah, "Holdout Validation for Comparison Classfication Naïve Bayes and KNN of Recipient Kartu Indonesia Pintar," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1125, no. 1, p. 012041, May 2021, doi: 10.1088/1757-899X/1125/1/012041.

[21] R. T. Nakatsu, "An Evaluation of Four Resampling Methods Used in Machine Learning Classification," *IEEE Intell. Syst.*, vol. 36, no. 3, pp. 51–57, May 2021, doi: 10.1109/MIS.2020.2978066.

[22] G. A. Syafarina and Z. Zaenuddin, "Implementasi Framework Streamlit Sebagai Prediksi Harga Jual Rumah Dengan Linear Regresi," *METIK J.*, vol. 7, no. 2, pp. 121–125, Dec. 2023, doi: 10.47002/metik.v7i2.608.

[23] M. Fadhilah and I. Ali, "Prediksi Jumlah Produksi Sablon Tahun Menggunakan Algoritma Regresi Linear di Nolbas SVNR".

[24] M. Akrom, S. Rustad, H. K. Dipojono, Prediction of Anti-Corrosion performance of new triazole derivatives via Machine learning, Comp. and Theoretical Chemistry, vol. 1236, pg. 114599, 2024.

[25] M. Obaseki and P. T. Elijah, "Application of Artificial Neural Network Model to Predict Corrosion Rates on Pipeline," vol. 3, no. 2, 2019.

[26] D. Leni, Y. P. Kusuma, R. Sumiati, Muchlisinalahuddin, and Adriansyah, "Perbandingan Alogaritma Machine Learning Untuk Prediksi Sifat Mekanik Pada Baja Paduan Rendah," *J. Rekayasa Mater. Manufaktur Dan Energi*, vol. 5, no. 2, Art. no. 2, Sep. 2022, doi: 10.30596/rmme.v5i2.11407.

[27] T. Richards, *Getting Started with Streamlit for Data Science: Create and deploy Streamlit web applications from scratch in Python*. Packt Publishing Ltd, 2021.

[28] V. N. G. Raju, K. P. Lakshmi, V. M. Jain, A. Kalidindi, and V. Padma, "Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification," in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Tirunelveli, India: IEEE, Aug. 2020, pp. 729–735. doi:

10.1109/ICSSIT48917.2020.9214160.
[29] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Appl. Soft Comput.*, vol. 97, p. 105524, Dec. 2020, doi: 10.1016/j.asoc.2019.105524.