Jurnal Ilmu Komputer dan Informasi (Journal of Computer Science and Information)
15/1 (2022), xx-xx. DOI: http://dx.doi.org/10.21609/jiki.v15i1.xxx

Inhibition Corrosion Website : Optimizing Ionization Energy Calculations with Linear and Nonlinear Algorithms

Dzaki Asari Surya Putra, Muhamad Akrom, and Gustina Alfa Trisnapradika

[1]Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang, 50131, Indonesia
[2]Research Group, Institution, Address, City, Zip Code, Country

E-mail: dzakiasari86@gmail.com

Abstract

In this research we implement Linear Algorithm and Nonlinear Algorithm on Inhibitor Corrosion Website, it is expected that scientists can provide predictions of compound ionization potential more efficiently and provide information about expired drugs in the context of corrosion inhibitors. Data collection, preprocessing, modeling, assessment, and implementation are all included in the standard machine learning implementation. Drug compounds are gathered into a dataset using pre-established descriptions. Additional data preparation is done to make the data ready for machine learning processing. To train the data, a modeling procedure is carried out using specified algorithms. The model is then evaluated using the provided descriptors to ascertain its accuracy and performance in target prediction. Lastly, an online application that realizes the model is made using Streamlit. As a more straightforward and effective method, we suggest utilizing the Corrosion Inhibitor Website to estimate the possible ionization energy of pharmaceutical chemicals that have expired. As a result, this application can help scientists determine the ionization energy of drug compounds more effectively, hence improving corrosion inhibition efficiency.

Keywords: *inhibition, corrosion, machine learning, linear, nonlinear*

1. Introduction

Corrosion is damage to metal or a chemical reaction process or electrochemical process between a metal and the surrounding environment which results in loss of metal quality [1], [2], [3]. Corrosion has a crucial role in the environmental context, and corrosion instruments can be adapted according to the conditions to which a metal is exposed, so that the consequences may be difficult to identify. The occurrence of corrosion is influenced by factors such as metal reactivity, degradation, the presence of air, humidity levels, as well as gas coefficients such as sulfur dioxide and carbon dioxide, and the presence of electrolytes [3]. Corrosion causes losses in various sectors including industrial assets, buildings and various other infrastructure [1], [4]. It costs around USD $2.5 trillion per year to combat corrosion problems [5]. So we need a corrosion inhibitor or inhibitor to follow up on existing corrosion problems [6].

Corrosion inhibitors have a very influential practical function, because they are often used to reduce metal losses during the production process and minimize the potential for material failure [2], [7], [8]. These two things can cause sudden disruptions in industrial operations, which can ultimately lead to additional costs. Studies show that drugs have an important role as inhibitors in various environments and at various concentration levels against various types of metals, such as zinc, copper, mild steel, aluminium, bronze, and carbon steel [3]. Drugs have the ability to form a protective layer on the surface of materials, preventing corrosive reactions between the material and the environment, thereby preventing or reducing the level of corrosion in certain materials [5]. Conventional methods for measuring inhibitor effectiveness are slow and costly [9]. Consequently, the use of machine learning offers a more convenient and rapid way to predict the corrosion inhibition efficiency of a compounds.

Machine learning is a branch of AI that allows systems to learn from data, recognize both linear and non-linear relationships, and improve performance without requiring complex programming from humans [10], [11], [12]. In the context of prediction and analysis of ionization potential energy, a Machine Learning (ML) approach is used to create a predictive model that

is able to estimate the level of corrosion in materials by considering the attributes that influence the event [6]. Based on previous research, it is proven that models with nonlinear algorithms provide the best performance compared to models with linear algorithms in predicting the ability to inhibit corrosion by Benzimidazole compounds based on $\boldsymbol{R^2}$, MAE, MSE and RMSE values [13].

In this research, we aim to develop an application that can help users estimate the ionization potential as a corrosion inhibitor in expired pharmaceutical compounds more efficiently. Previous research has compared the performance of two machine learning algorithms for predicting the efficiency of pyridazine corrosion inhibitors. The research results show that the Decision Tree Regressor algorithm is more accurate than the multilinear regression algorithm. This is indicated by the higher R2, MAE and RMSE values in the Decision Tree Regressor Algorithm [14]. Other research on the application of nonlinear Machine Learning algorithms on expired corrosion inhibitor drugs datasets using a Gradient Boosting Regressor with a polynomial function provides good performance with an R2 value of 0.9998 for training data and 0.9999 for testing data so that it can provide good predictions of potential ionization energy [6]. By implementing Linear Algorithms and Nonlinear Algorithms on the Inhibitor Corrosion Website, it is hoped scientist can provide predictions of the ionization potential of compounds more efficiently and provide information about expired drugs in the context of corrosion inhibitors [15].

2. Method

The implementation of machine learning generally includes data collection, data preprocessing, modeling, evaluation and implementation. Drug compound dataset is collected based on predetermined descriptors. To get the data ready for more machine learning processing, data preparation is next done. Using preset algorithms, the modeling process is run to train data. Subsequently, the model undergoes an assessment to determine its performance and accuracy in target prediction using available descriptors. Lastly, streamlit is used to create an online application that embodies the model. The scheme of the research carried out is illustrated in Figure 1.

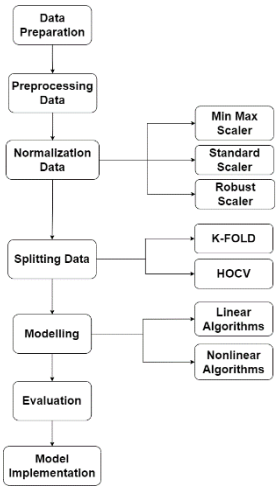

Fig. 1. Research Metodology

2.1. Data Preparation

In this study, we used a dataset containing 250 expired drug compounds with 10 molecular descriptor features [6], [16]. The chemical reaction in the inhibitor molecule greatly influences corrosion inhibition. Descriptors that are generally used to evaluate the correlation of corrosion inhibitor effectiveness are HOMO, LUMO energy, energy (ΔE), ionization potential (I), electron affinity (A), global hardness (η), global softness (σ), electronegativity (χ , dipole moment (μ), electrophilicity (ω), fraction of electrons transferred (ΔN), total energy (TE), and also other descriptors such as inhibitor concentration, temperature, polarizability, and other descriptors determined through experiments [5].

2.2. Preprocessing Data

Preprocessing is done to separate data in order to identify missing values and outliers in the data [6]. In order to further assess the sensitivity of the machine learning model, the data is normalized so that each feature has a similar scale [17], [18]. In this study we used three normalization techniques, namely MinMax Scaler, Standard Scaler, and Robust Scaler.

MinMaxScaler normalization is a type of feature scaling that rescales the values of each feature in a dataset to a common range. This is done by subtracting the minimum value of each feature from all values in the feature and then dividing by the range of values in the feature [19].

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

where Xmin = minimum value in X Feature and Xmax = maximum value in X Feature. MinMax

Scaler works by transforming the data so that all values fall within a specified range, typically between 0 and 1.

Standard Scaler assumes your data within each feature follows a normal distribution. It then scales those features so that the distribution after scaling is centered around 0 with a standard deviation of 1. To achieve this, it calculates the mean and standard deviation for each feature and then applies these values to scale the feature according to specific formulas [19] :

$$Zscaled = \frac{(X-\mu)}{\sigma} \qquad (2)$$

where $\mu = Mean\ and\ \sigma = Standard\ Deviation$ While the Min-Max scaler operates on the min and max values, the RobustScaler employs a more outlier-resistant approach by utilizing the interquartile range (IQR). This robust scaler first centers the data on the median and then scales it according to the IQR, defined as the difference between the 25th and 75th percentiles [19].

$$RS(Xi) = \frac{(Xi-median(X))}{IQR_{1,3}(X)} \quad (3)$$

where $IQR_{1,3}(X)$ = Interquartile distance between Q1 (25^{th}) and Q3 (75^{th}) percentiles.

2.3. Splitting Data

Data splitting is done to identify bias and variance in the data. [20]. We used HoldOut Cross Validation (HOCV) and K-Fold Cross Validation (KFCV) in this study. Holdout validation utilizes a split-data approach where each data point has the chance to be part of the training set and the testing set, essentially undergoing the accuracy test twice [21]. Whereas K-Fold Cross Validation (KFCV) avoid overfitting and get a better sense of how well a model generalizes, k-fold cross-validation iterates k times. Each time, it takes one chunk of the data as the validation set and trains the model on the remaining chunks. By averaging the validation errors across all k runs, it obtains a more robust estimate of the model's performance [22].

2.4. Modelling

The Corrosion Inhibitor Website requires a machine learning model to accurately predict the ionization energy potential of a compound. This would make it easier for scientists to select the most effective corrosion inhibitors for a given application. In the realm of model building, we employ a repertoire of both linear and nonlinear algorithms to harness the diverse range of data complexities [13].

Linear Algorithms will be used to understand the relationship between predictor (independent variable) and target (dependent variable). The model will be checked to ensure that the basic assumptions of linear algorithm are met [13], [14], [23], [24]. The nonlinear algorithms used in this study include Linear Regression, Ridge, Lasso, and Elastic Net.

Nonlinear algorithms are a type of algorithm that expresses the relationship between dependent (target) and independent (predictor) variables in a nonlinear form. Models with nonlinear algorithms tend to have better accuracy in estimating the model using algorithm iterations [6], [15], [25]. The nonlinear algorithms used in this research is XGB Regressor, CatBoost Regressor, Light GBM Regressor, AdaBoost Regressor, NuSVR, SVR, Decision Tree Regressor, Gradient Boosting Regressor, and Random Forest Regressor.

2.5. Evaluation

To improve the model's performance in providing accurate predictions on the Korosi Inhibitor Website, we employ regression metrics such as RMSE R^2 score. RMSE (Root Mean Squared Error) is a metric used to quantify the accuracy of predictions by measuring the average squared difference between predicted values and actual values, serta R^2 is a measure of how well the predicted values fit the actual values. It has a range of values from 0 to 1 [14]. A machine learning model with the minimum RMSE values and R^2 score closest to 1 is the best machine learning model [13], [26], [27].

2.6. Implementation

After completing the modeling, the next step is to apply the model through a web application using Streamlit, making it easier for scientists to consider the potential of ionization energy in corrosion inhibition. Streamlit is a Python-based web application framework that is open source and has advantages in speed and flexibility of development. In general, Streamlit is used for analysis, creation of complex interactive applications, and application of machine learning models [10], [28]. In designing a machine learning user interface, researchers created two options for predicting ionization potential. These options include prediction based on a single input and prediction based on a CSV file containing descriptor features of drug compounds.

3. Result and Discussion

The RMSE values for each combination of linear and nonlinear algorithms were determined through training data that utilized a combination of normalization and data splitting for each algorithm. The RMSE values for each combination are presented in Table 1.

Table 1. RMSE and R2 scores of linear and nonlinear algorithm combination in data training

Splitting Data	Algorithm	Scaler	RMSE	R^2 Score
HOCV 0.2	LR	MinMax	6.1577	0.0909
		Standard	6.1547	0.0918
		Robust	6.1547	0.0918
		No Scale	6.1547	0.0918
	RR	MinMax	6.2041	0.0772
		Standard	6.1558	0.0915
		Robust	6.1565	0.0913
		No Scale	6.1702	0.0872
	EN	MinMax	6.4585	0.0
		Standard	6.3277	0.0401
		Robust	6.3031	0.0475
		No Scale	6.2611	0.0602
	Lasso	MinMax	6.4585	0.0
		Standard	6.3755	0.0255
		Robust	6.3559	0.0315
		No Scale	6.2985	0.0489
	XGB	MinMax	0.0006	0.9999
		Standard	0.0006	0.9999
		Robust	0.0006	0.9999
		No Scale	0.0006	0.9999
	GBR	MinMax	0.0175	0.9932
		Standard	0.4298	0.9955
		Robust	0.4298	0.9955
		No Scale	0.4298	0.9955
	CatBoost	MinMax	0.0350	0.9999
		Standard	0.0350	0.9999
		Robust	0.0350	0.9999
		No Scale	0.0350	0.9999
	LGBM	MinMax	5.4030	0.3001
		Standard	5.2520	0.3387
		Robust	5.4136	0.2973
		No Scale	5.4034	0.3000
	AdaBoost	MinMax	2.6514	0.8314
		Standard	2.5271	0.8468
		Robust	2.6394	0.8329
		No Scale	2.5550	0.8434
	NuSVR	MinMax	6.2645	0.0591
		Standard	6.1608	0.0900
		Robust	6.2054	0.0768
		No Scale	6.4134	0.0139
	SVR	MinMax	6.5965	-0.0431
		Standard	6.5075	-0.0152
		Robust	6.5648	-0.0331
		No Scale	6.8044	-0.1099
	DT	MinMax	0.0	1.0
		Standard	0.0	1.0
		Robust	0.0	1.0
		No Scale	0.0	1.0
	RF	MinMax	2.8411	0.8064
		Standard	2.6154	0.8360
		Robust	2.6757	0.8283
		No Scale	2.8489	0.8054
HOCV 0.3	LR	MinMax	6.4660	0.0815
		Standard	6.4421	0.0883
		Robust	6.4421	0.0883
		No Scale	6.4421	0.0883
	RR	MinMax	6.4999	0.0719
		Standard	6.4445	0.0876
		Robust	6.4453	0.0874
		No Scale	6.4674	0.0811
	Elastic	MinMax	6.7471	0.0
		Standard	6.6128	0.0393
		Robust	6.5924	0.0453
		No Scale	6.5415	0.0
	Lasso	MinMax	6.7471	0.0
		Standard	6.6833	0.0188
		Robust	6.6492	0.0287
		No Scale	6.5759	0.0500
	XGB	MinMax	0.0006	0.9999
		Standard	0.0006	0.9999
		Robust	0.0006	0.9999
		No Scale	0.0006	0.9999
	GBR	MinMax	0.0015	0.9952
		Standard	0.4503	0.9955
		Robust	0.4503	0.9955
		No Scale	0.4503	0.9955
	CatBoost	MinMax	0.0574	0.9999
		Standard	0.0574	0.9999
		Robust	0.0574	0.9999
		No Scale	0.0574	0.9999
	LGBM	MinMax	5.6471	0.2994
		Standard	5.5842	0.3150
		Robust	5.5772	0.3166
		No Scale	5.5817	0.3156
	AdaBoost	MinMax	2.5963	0.8519
		Standard	2.5525	0.8568
		Robust	2.6262	0.8484
		No Scale	2.2659	0.8872
	NuSVR	MinMax	6.5633	0.0537
		Standard	6.4767	0.0785
		Robust	6.5329	0.0624
		No Scale	6.7306	0.0048
	SVR	MinMax	7.0282	-0.0850
		Standard	6.8687	-0.0363
		Robust	7.0230	-0.0834
		No Scale	7.3534	-0.1878
	DT	MinMax	0.0	1.0
		Standard	0.0	1.0
		Robust	0.0	1.0
		No Scale	0.0	1.0
	RF	MinMax	2.9439	0.8096
		Standard	2.9310	0.8112
		Robust	2.9248	0.8120
		No Scale	2.7077	0.8389
HOCV 0.4	LR	MinMax	6.4390	0.1365
		Standard	6.4390	0.1365
		Robust	6.4390	0.1365
		No Scale	6.4390	0.1365
	RR	MinMax	6.6338	0.0835
		Standard	6.4559	0.1320
		Robust	6.4605	0.1307
		No Scale	6.5331	0.1111
	Elastic	MinMax	6.9295	0.0
		Standard	6.7531	0.0502
		Robust	6.7220	0.0589
		No Scale	6.6423	0.0811
	Lasso	MinMax	6.9295	0.0
		Standard	6.8184	0.0318
		Robust	6.7794	0.0428
		No Scale	6.6692	0.0737
	XGB	MinMax	0.0007	0.9999
		Standard	0.0007	0.9999
		Robust	0.0007	0.9999
		No Scale	0.0007	0.9999
	GBR	MinMax	0.0076	0.9988
		Standard	0.2982	0.9981

		Robust	0.2982	0.9981			No Scale	7.0079	-0.1930
		No Scale	0.2982	0.9981		DT	MinMax	10.1367	-1.4502
	CatBoost	MinMax	0.0351	0.9999			Standard	10.1054	-1.3669
		Standard	0.0351	0.9999			Robust	10.0017	-1.3225
		Robust	0.0351	0.9999			No Scale	10.3247	-1.7642
		No Scale	0.0351	0.9999		RF	MinMax	7.6146	-0.4639
	LGBM	MinMax	6.5172	0.1154			Standard	7.7279	-0.5039
		Standard	6.4606	0.1307			Robust	7.7796	-0.4795
		Robust	6.4345	0.1377			No Scale	7.7424	-0.4225
		No Scale	6.4073	0.1450	K-Fold	LR	MinMax	7.6483	-0.5886
	AdaBoost	MinMax	2.2873	0.8910	5		Standard	7.6818	-0.6028
		Standard	2.1265	0.9058			Robust	7.6818	-0.6028
		Robust	2.2353	0.8959			No Scale	7.6818	-0.6028
		No Scale	2.4334	0.8766		RR	MinMax	6.7554	-0.2220
	NuSVR	MinMax	6.7811	0.0423			Standard	7.4837	-0.5132
		Standard	6.6824	0.0700			Robust	7.4717	-0.5076
		Robust	6.7323	0.0560			No Scale	7.5103	-0.5140
		No Scale	6.9150	0.0041		Elastic	MinMax	6.5255	-0.1378
	SVR	MinMax	7.1903	-0.0766			Standard	6.5021	-0.1317
		Standard	7.0849	-0.0453			Robust	6.5752	-0.1529
		Robust	7.1650	-0.0691			No Scale	7.0883	-0.3599
		No Scale	7.4140	-0.1447		Lasso	MinMax	6.5255	-0.1378
	DT	MinMax	0.0	1.0			Standard	6.4643	-0.1149
		Standard	0.0	1.0			Robust	6.5441	-0.1393
		Robust	0.0	1.0			No Scale	6.8744	-0.2792
		No Scale	0.0	1.0		XGB	MinMax	8.7423	-1.5601
	RF	MinMax	2.9889	0.8139			Standard	8.7423	-1.5601
		Standard	3.1180	0.7975			Robust	8.7423	-1.5601
		Robust	2.9578	0.8177			No Scale	8.7423	-1.5601
		No Scale	2.9680	0.8165		GBR	MinMax	8.4334	-1.0725
K-Fold	LR	MinMax	7.6524	-0.4187			Standard	8.3502	-1.0530
3		Standard	7.5739	-0.3922			Robust	8.3207	-1.0548
		Robust	7.5739	-0.3922			No Scale	8.4164	-1.1077
		No Scale	7.5739	-0.3922		CatBoost	MinMax	7.6836	-0.6746
	RR	MinMax	6.7410	-0.1036			Standard	7.6841	-0.6748
		Standard	7.3132	-0.2985			Robust	7.6877	-0.6758
		Robust	7.3285	-0.3054			No Scale	7.6826	-0.6740
		No Scale	7.3425	-0.3123		LGBM	MinMax	6.8325	-0.2826
	Elastic	MinMax	6.6084	-0.0631			Standard	6.9033	-0.3281
		Standard	6.5547	-0.0448			Robust	6.9523	-0.3663
		Robust	6.6340	-0.0710			No Scale	6.8660	-0.3226
		No Scale	7.0405	-0.2107		AdaBoost	MinMax	7.8732	-0.6888
	Lasso	MinMax	6.6084	-0.0631			Standard	7.9119	-0.7677
		Standard	6.5324	-0.0374			Robust	7.6440	-0.6901
		Robust	6.6130	-0.0648			No Scale	7.8396	-0.7129
		No Scale	6.8470	-0.1433		NuSVR	MinMax	6.4106	-0.0919
	XGB	MinMax	9.4004	-1.2421			Standard	6.3835	-0.0837
		Standard	9.4004	-1.2421			Robust	6.4032	-0.0912
		Robust	9.4004	-1.2421			No Scale	6.4970	-0.1187
		No Scale	9.4004	-1.2421		SVR	MinMax	6.7603	-0.2059
	GBR	MinMax	9.1448	-1.0388			Standard	6.7615	-0.2046
		Standard	9.1079	-1.0708			Robust	6.7727	-0.2122
		Robust	9.1913	-1.0168			No Scale	6.9038	-0.2572
		No Scale	9.2359	-1.0321		DT	MinMax	10.3422	-2.4454
	CatBoost	MinMax	7.8091	-0.5019			Standard	10.3338	-2.3362
		Standard	7.8082	-0.5016			Robust	9.9191	-2.1388
		Robust	7.8095	-0.5021			No Scale	10.1510	-2.3664
		No Scale	7.8093	-0.5020		RF	MinMax	7.2984	-0.5330
	LGBM	MinMax	6.6435	-0.0779			Standard	7.3387	-0.5992
		Standard	6.4920	-0.0273			Robust	7.2722	-0.5402
		Robust	6.6108	-0.0650			No Scale	7.3914	-0.4988
		No Scale	6.5413	-0.0452	K-Fold	LR	MinMax	7.2414	-0.6218
	AdaBoost	MinMax	8.0987	-0.5186	10		Standard	7.2492	-0.6232
		Standard	7.9918	-0.7163			Robust	7.3078	-0.6581
		Robust	8.1492	-0.6001			No Scale	7.3078	-0.6581
		No Scale	8.2316	-0.6171		RR	MinMax	6.5265	-0.2760
	NuSVR	MinMax	6.5428	-0.0396			Standard	7.0970	-0.5387
		Standard	6.5129	-0.0304			Robust	7.0753	-0.5256
		Robust	6.5398	-0.0388			No Scale	6.9957	-0.4795
		No Scale	6.5873	-0.0563		Elastic	MinMax	6.3919	-0.2303
	SVR	MinMax	6.8661	-0.1480			Standard	6.2997	-0.1881
		Standard	6.8237	-0.1322			Robust	6.3488	-0.2056
		Robust	6.8732	-0.1493			No Scale	6.6146	-0.3417

Lasso	MinMax	6.3919	-0.2303
	Standard	6.2868	-0.1808
	Robust	6.3245	-0.1973
	No Scale	6.5302	-0.3047
XGB	MinMax	7.9251	-1.0394
	Standard	7.9251	-1.0394
	Robust	7.9251	-1.0394
	No Scale	7.9251	-1.0394
GBR	MinMax	8.1343	-1.0519
	Standard	8.0733	-1.0712
	Robust	8.0933	-1.0468
	No Scale	8.0701	-1.0590
CatBoost	MinMax	7.3660	-0.6552
	Standard	7.3660	-0.6552
	Robust	7.3661	-0.6558
	No Scale	7.3638	-0.6544
LGBM	MinMax	6.7545	-0.3653
	Standard	6.6784	-0.3576
	Robust	6.7241	-0.3953
	No Scale	6.6437	-0.3413
AdaBoost	MinMax	7.2598	-0.6548
	Standard	7.2314	-0.4990
	Robust	7.2666	-0.4150
	No Scale	7.5674	-0.5825
NuSVR	MinMax	6.2564	-0.1671
	Standard	6.2534	-0.1643
	Robust	6.2545	-0.1683
	No Scale	6.3840	-0.2134
SVR	MinMax	6.5531	-0.2585
	Standard	6.5393	-0.2483
	Robust	6.5476	-0.2552
	No Scale	6.7427	-0.3344
DT	MinMax	9.6368	-2.4186
	Standard	9.0377	-1.9295
	Robust	9.1795	-1.9059
	No Scale	9.4755	-2.3403
RF	MinMax	7.1252	-0.5071
	Standard	6.9030	-0.5305
	Robust	6.9632	-0.5107
	No Scale	6.9710	-0.5745

Based on the conducted testing, models exhibiting favorable R^2 and RMSE scores are predominantly associated with non-linear algorithm models. As a result, these models are more recommended for predicting the ionization energy potential in the case study of corrosion inhibitor compounds [6], [13].

The non-optimal RMSE and R^2 scores in Table 1 indicate that the linear algorithm model tends to forecast ionization energy with an inadequate level of accuracy. Non-linear algorithms, on the other hand, may estimate a drugs compound's potential ionization energy quite well.

Subsequently, the models generated from training data using a combination of different algorithms, data splitting, and data normalization are implemented into a web-based application called Streamlit. This is illustrated in Figure 2 and Figure 3.

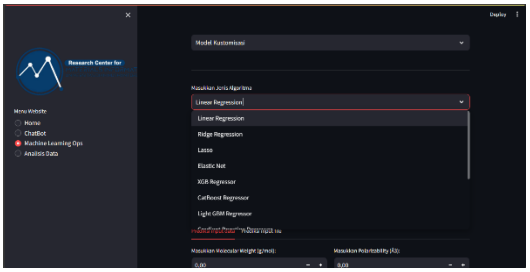

Fig. 2. The page predicts the ionization energy on the Corrosion Inhibitor Website, which implements various algorithms, including both linear and nonlinear approaches.

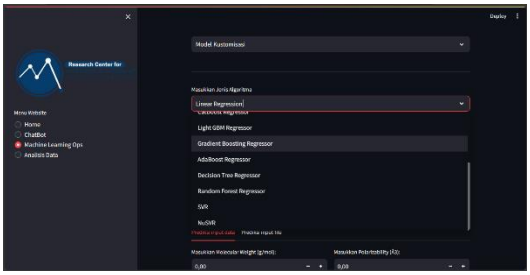

Fig. 3. The page predicts the ionization energy on the Corrosion Inhibitor Website, which implements various algorithms, including both linear and nonlinear approaches.

The ionization energy prediction page offers a variety of linear and non-linear algorithms that can be chosen for prediction. This provides space for scientists to explore the application of machine learning in predicting ionization energy in corrosion inhibitor drug compounds. Additionally, the option for normalizing training data is also provided, as illustrated in Figure 4.

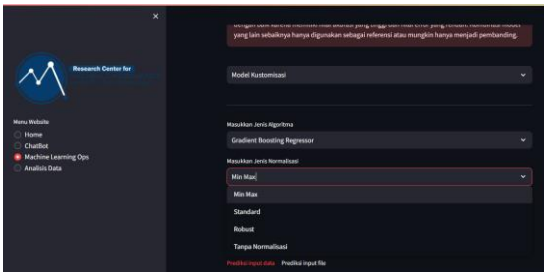

Fig. 4. Options for data normalization on the ionization energy prediction page on the Corrosion Inhibitor Website.

Data normalization used to improve the sensitivity analysis of the machine learning model and bringing all features to a common scale [19], [29]. Furthermore, we also provide the option to choose data splitting using either HOCV or K-Fold, as depicted in Figure 5.

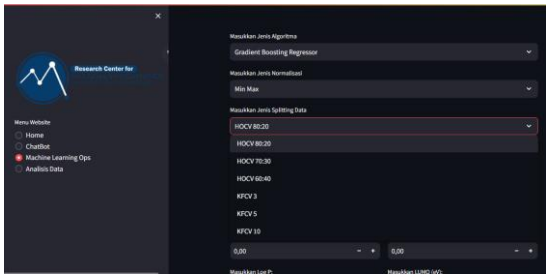

Fig. 5. HOCV and K-Fold options for splitting data in ML model.

HOCV and K-Fold are employed for data partitioning to discern potential bias and variance within the dataset [20]. To facilitate the prediction process, we also provide options for the best models, as seen in Figure 6, for your convenience.

Fig. 6. Best Model Algorithm option to predict ionization energy.

Utilizing this best model option, scientists can efficiently attain the ionization potential of compounds. This model has been pre-trained using the Gradient Boosting Regressor algorithm with a polynomial function, demonstrating excellent performance with an R^2 score of 0.9998 for training data and 0.9999 for testing data [6]. Afterward, scientists can leverage the capabilities of machine learning to forecast the ionization energy of compounds based on their existing features. This is illustrated in Figure 7.

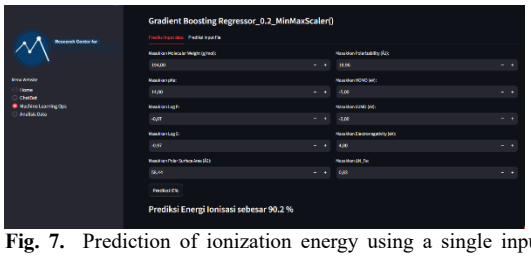

Fig. 7. Prediction of ionization energy using a single input dataset.

To ascertain the ionization energy potential of a compound, scientists simply need to input characteristic data of the compound they wish to predict. Beyond using a single input data, there is an option to upload a CSV file containing data of the compounds to be predicted, considering the relevant features. Refer to Figure 8 for details.

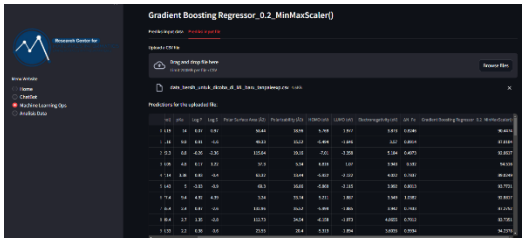

Fig. 8. Prediction of ionization energy using csv file input.

After uploading the data, a dataframe view will appear, displaying the CSV file data that has been uploaded along with an additional column showing the predicted ionization energy generated from machine learning, as illustrated in Figure 8.

4. Conclusion

To predict the ionization energy potential of expired pharmaceutical compounds, we propose the use of the Corrosion Inhibitor Website as a more straightforward and efficient option. In the Streamlit web-based application, machine learning has been successful in predicting potential ionization energy. Therefore, this implementation can provide valuable support for scientists in calculating the ionization energy of drugs compounds in order to improve efficiency in corrosion inhibition. It is crucial to consider alternative algorithms or methods when implementing a machine learning approach for predicting the ionization energy of drugs compounds to enhance accuracy and prediction efficiency.

Acknowledgement

Thank you Allah, SWT.

References

[1] P. D. N. H. S. M.T S. T., S. M.T S. T., And S. H. M.T S. T., *Pengantar Inhibitor Korosi Alami*. Deepublish, 2021.

[2] "Corrosion Inhibitors. A Review," *Int. J. Corros. Scale Inhib.*, Vol. 10, No. 1, Mar. 2021, Doi: 10.17675/2305-6894-2021-10-1-3.

[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] Y. Stiadi, S. Arief, H. Aziz, M. Efdi, And E. Emriadi, "Inhibisi Korosi Baja Ringan Menggunakan Bahan Alami Dalam Medium Asam Klorida: Review," *J. Ris. Kim.*, Vol. 10, No. 1, Pp. 51–65, Mar. 2019, Doi: 10.25077/Jrk.V12i2.321.

[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] D. Leni, Y. P. Kusuma, M. Muchlisinalahuddin, F. Earnerstly, R. Muharni, And R. Sumiati, "Perancangan Metode Machine Learning Berbasis Web Untuk Prediksi Sifat Mekanik Aluminium," *J. Rekayasa Mesin*, Vol. 14, No. 2, Pp. 611–626, Aug. 2023, Doi: 10.21776/Jrm.V14i2.1370.

[11] Prima Fierza Saputra, D. Leni, And F. Earnestly, "Prediksi Sifat Mekanik Aluminium Berdasarkan Unsur Kimia Paduan Menggunakan Model Machine Learning," *J. Surya Tek.*, Vol. 10, No. 2, Pp. 799–804, Dec. 2023, Doi: 10.37859/Jst.V10i2.5809.

[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] 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.

[17] 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.

[18] S. Rahayu And J. J. Purnama, "Klasifikasi Konsumsi Energi Industri Baja Menggunakan Teknik Data Mining," *J. Teknoinfo*, Vol. 16, No. 2, P. 395, Jul. 2022, Doi: 10.33365/Jti.V16i2.1984.

[19] 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.

[20] M. Akrom And T. Sutojo, "Investigasi Model Machine Learning Berbasis QSPR Pada Inhibitor Korosi Pirimidin," *Eksergi*, Vol. 20, No. 2, P. 107, Jul. 2023, Doi: 10.31315/E.V20i2.9864.

[21] 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.

[22] 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.

[23] 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.

[24] M. Fadhilah And I. Ali, "Prediksi Jumlah Produksi Sablon Tahun Menggunakan Algoritma Regresi Linear Di Nolbas SVNR".

[25] W. Mujahid And M. A. Tiro, "Pemodelan Laju Inflasi Dengan Menggunakan Regresi Non-Linear Berbasis Algoritma Genetika".

[26] M. Obaseki And P. T. Elijah, "Application Of Artificial Neural Network Model To Predict Corrosion Rates On Pipeline," Vol. 3, No. 2, 2019.

[27] 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.

[28] T. Richards, *Getting Started With Streamlit For Data Science: Create And Deploy Streamlit Web Applications From Scratch In Python*. Packt Publishing Ltd, 2021.

[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.