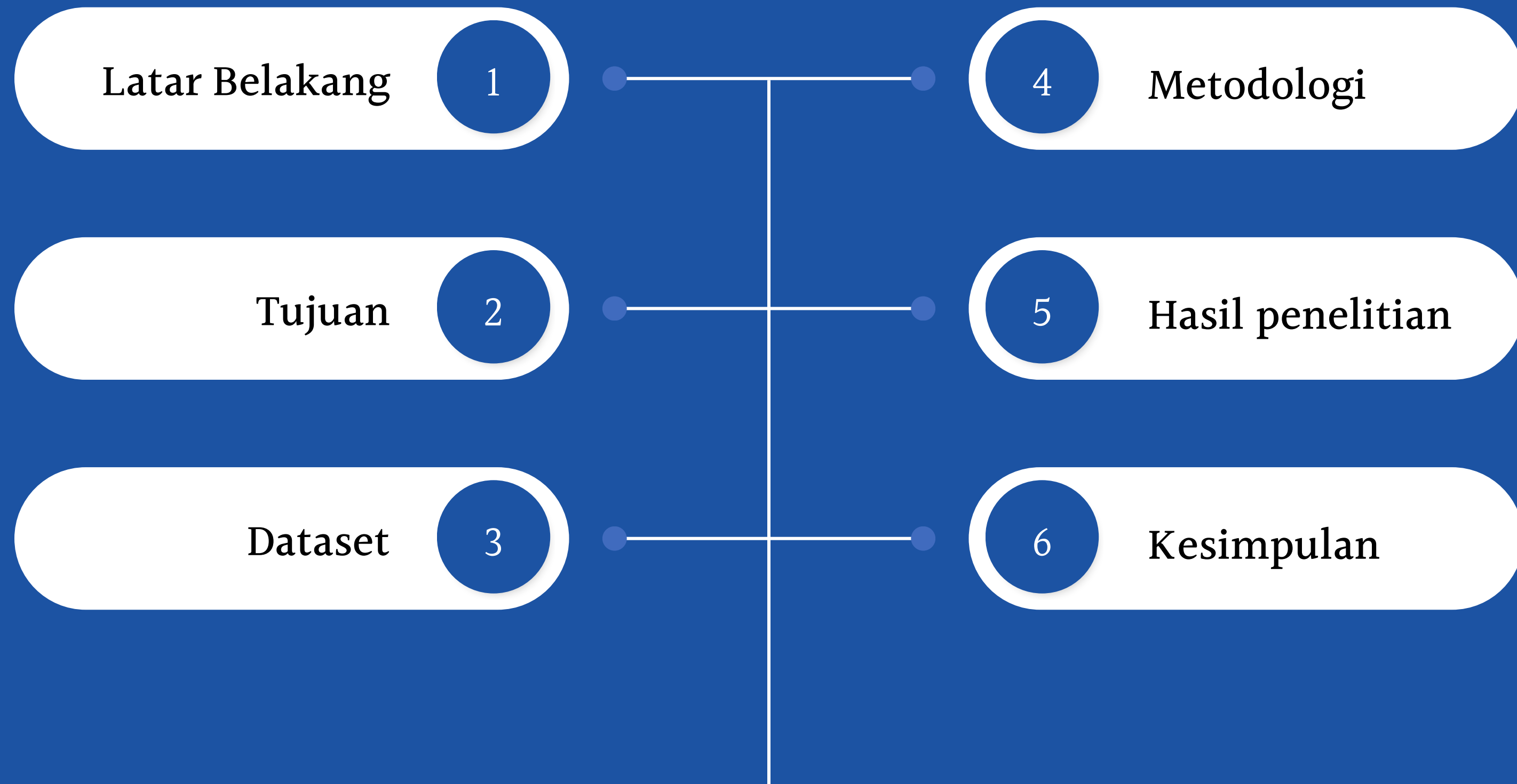


Perbandingan Algoritma **Deep Learning** dan **Machine Learning** untuk prediksi senyawa **N-Heterocyclic** sebagai **Corrosion Inhibitor**

Rizky Syah G.



Pembahasan



Latar Belakang

Korosi

Korosi adalah proses degradasi material, biasanya logam, akibat reaksi kimia dengan lingkungannya. Proses ini mengakibatkan kerusakan struktural dan penurunan kualitas material. Korosi dapat terjadi melalui berbagai mekanisme, termasuk reaksi elektrokimia antara logam dan oksigen atau air.

N-Heterocyclic

N-Heterocyclic sering digunakan karena kemampuannya untuk berinteraksi dengan permukaan logam dan membentuk lapisan pelindung yang mencegah reaksi korosi. Senyawa ini dapat efektif dalam berbagai lingkungan korosif dan digunakan dalam berbagai aplikasi industri.

ML vs DL

ML menggunakan berbagai teknik statistik untuk menemukan pola dalam data. Deep Learning (DL) adalah subbidang dari machine learning yang menggunakan jaringan saraf tiruan dengan banyak lapisan untuk memodelkan hubungan yang kompleks dalam data. DL sangat efektif dalam menangani data yang tidak terstruktur, seperti gambar, suara, dan teks.

Tujuan

Membandingkan performa algoritma deep learning dan machine learning dalam memprediksi efikasi senyawa N-Heterocyclic sebagai corrosion inhibitor.

Mencari algoritma dengan akurasi tertinggi pada setiap algoritma untuk mengetahui prediksi penghambatan korosi pada senyawa n-heterocyclic

Meningkatkan pemahaman terhadap masing masing hasil prediksi untuk memahami kekuatan dan kelemahan masing-masing algoritma.

Dataset

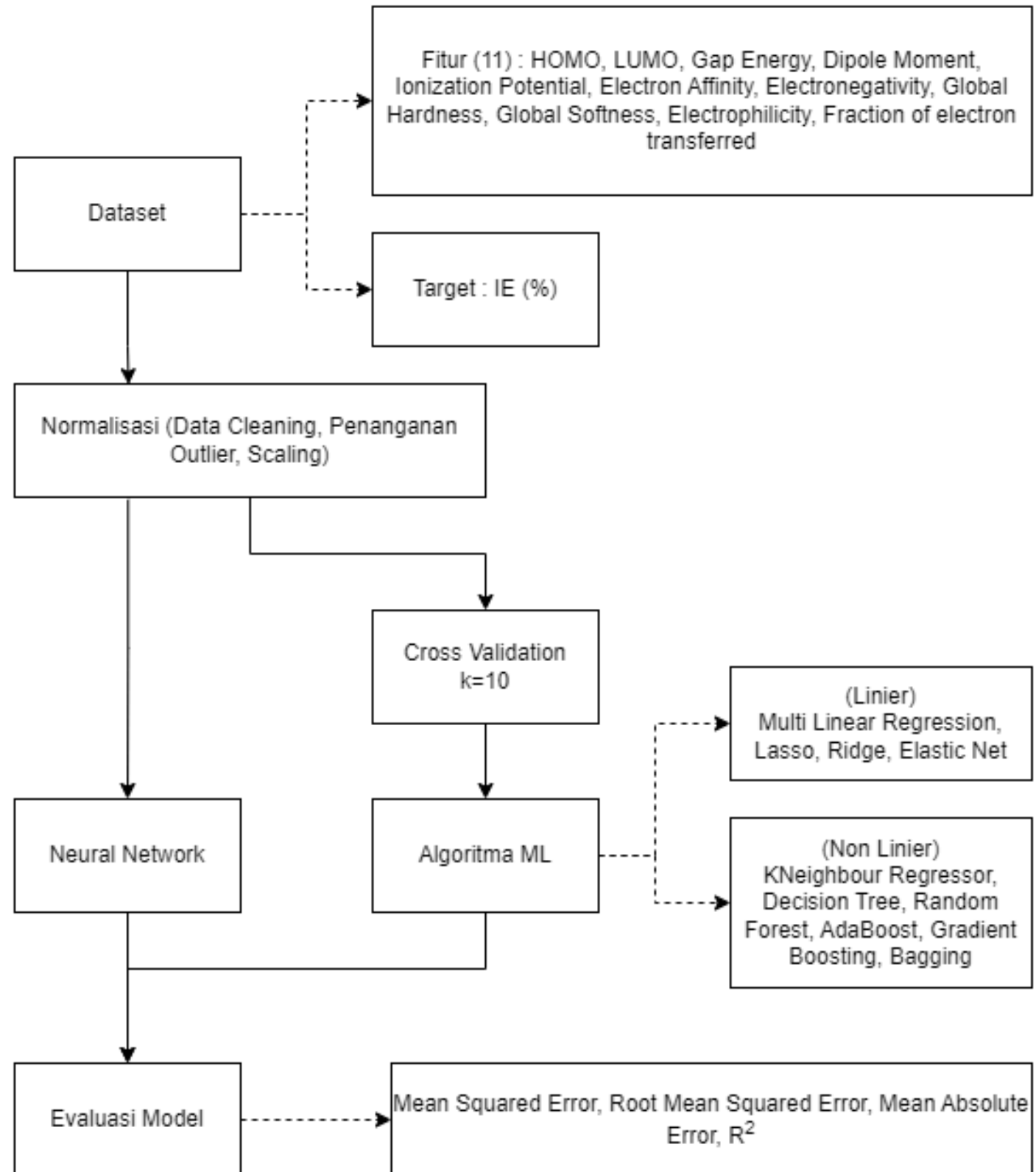
HOMO	LUMO	Gap Energy	Dipole Moment	Ionization Potential	Electron Affinity	Electronegativity	Global Hardness	Global Softness	Electrophilicity	Fraction of electron transferred	IE (%)
-6,12	-1,852	4,268	3,425	6,996	1,446	4,221	2,775	0,36	3,211	0,501	13
-6,034	-1,788	4,246	2,976	6,749	1,404	4,076	2,673	0,374	3,109	0,547	19
-6,048	-1,809	4,239	3,872	7,561	1,427	4,494	3,067	0,326	3,292	0,409	14
-6,091	-1,71	4,381	4,252	6,928	1,321	4,125	2,804	0,357	3,034	0,513	24
-5,954	-1,728	4,226	2,437	6,525	1,366	3,946	2,579	0,388	3,018	0,592	26
-5,931	-1,596	4,335	3,239	8,097	1,244	4,67	3,427	0,292	3,183	0,34	38
-5,968	-1,84	4,128	3,246	6,301	1,524	3,912	2,388	0,419	3,205	0,646	88
-5,973	-1,861	4,112	3,531	6,3	1,545	3,922	2,378	0,421	3,235	0,647	93
-6,008	-1,862	4,146	3,897	6,334	1,584	3,959	2,375	0,421	3,3	0,64	72
-6,081	-1,496	4,585	4,873	6,6	1,125	3,862	2,738	0,365	2,725	0,573	28
-5,937	-1,798	4,139	5,17	6,448	1,407	3,927	2,521	0,397	3,059	0,609	73
-5,843	-1,831	4,012	5,33	6,363	1,429	3,896	2,467	0,405	3,077	0,629	21
-6,127	-3,17	2,957	2,695	6,73	2,781	4,756	1,975	0,506	5,727	0,568	77
-6,262	-3,374	2,888	2,747	6,818	2,956	4,887	1,931	0,518	6,183	0,547	49
-6,48	-2,062	4,418	5,102	7,018	1,63	4,324	2,694	0,371	3,471	0,497	61
-6,345	-2,064	4,281	5,041	6,968	1,607	4,287	2,681	0,373	3,429	0,506	32
-5,028	-1,204	3,824	7,279	5,495	0,882	3,188	2,306	0,434	2,204	0,826	61
-5,771	-1,731	4,039	1,205	6,255	1,347	3,801	2,454	0,408	2,944	0,652	73
-6,001	-1,739	4,262	3,823	6,705	1,334	4,02	2,686	0,372	3,008	0,555	23
-5,961	-1,759	4,202	2,957	6,71	1,361	4,035	2,675	0,374	3,044	0,554	16

Rows : 218

Fitur (11) : HOMO, LUMO, Gap Energy, Dipole Moment, Ionization Potential, Electron Affinity, Electronegativity, Global Hardness, Global Softness, Electrophilicity, Fraction of electron transferred

Target : IE (%)

Metodologi

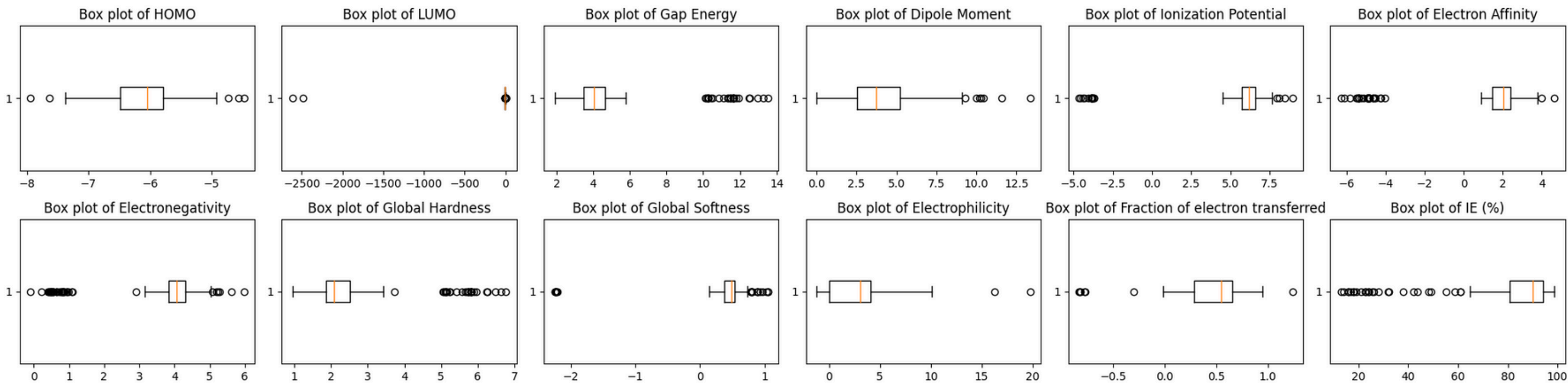


Model: "sequential_20"

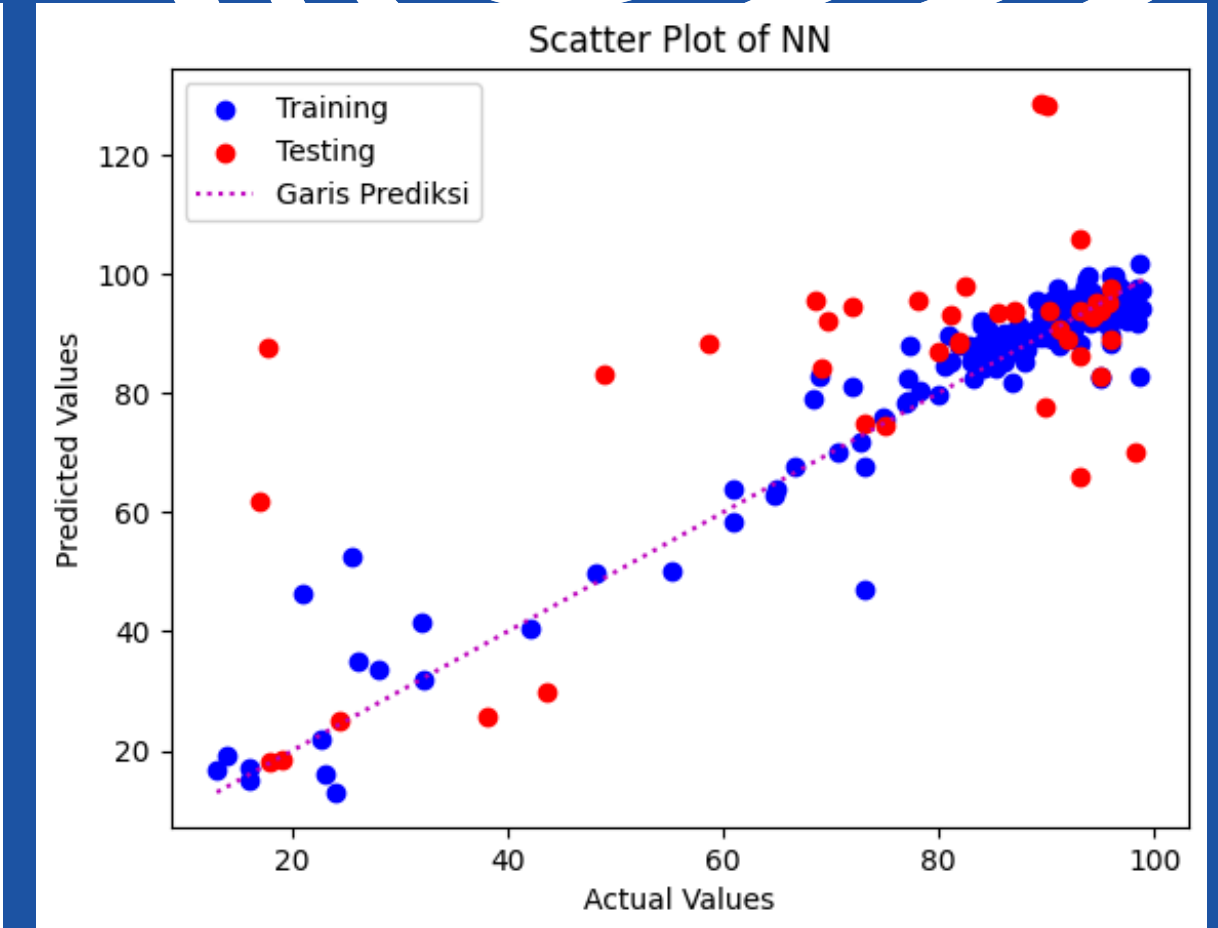
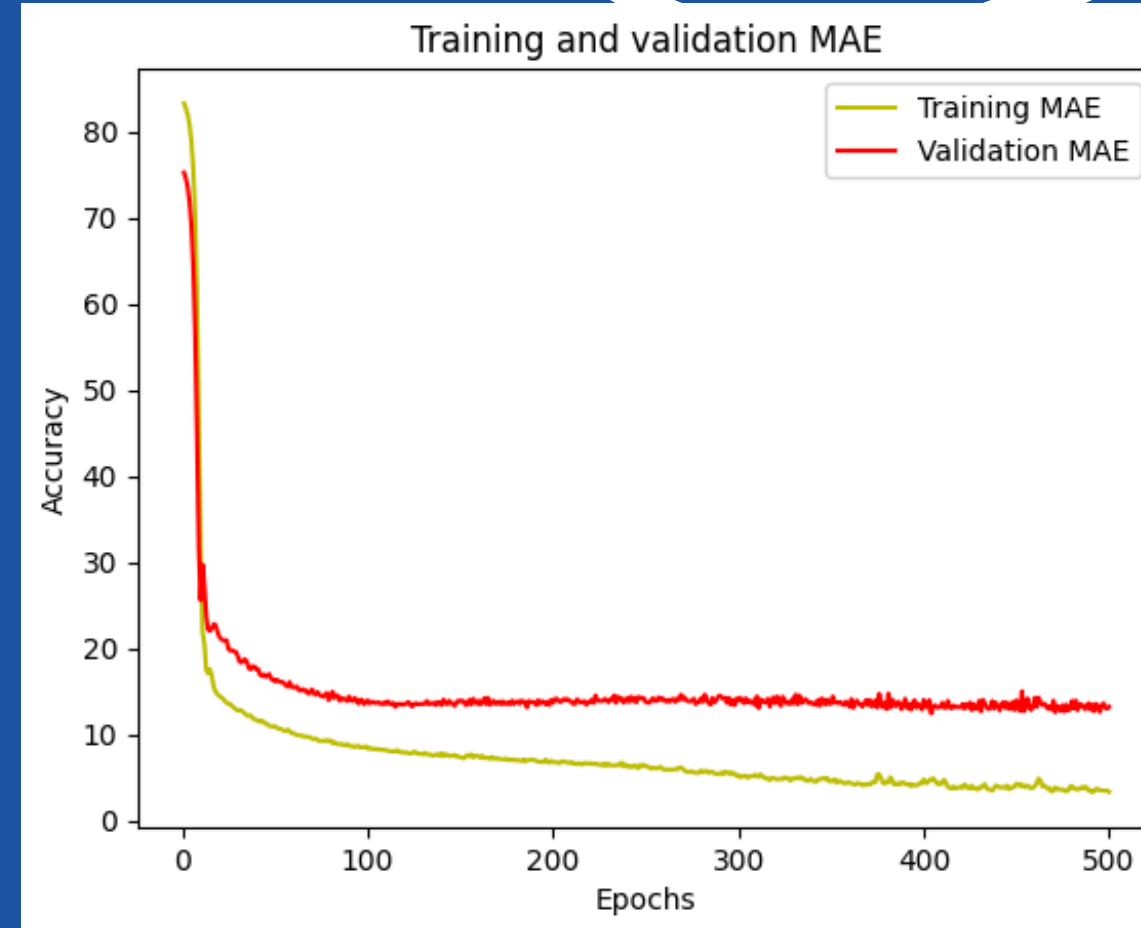
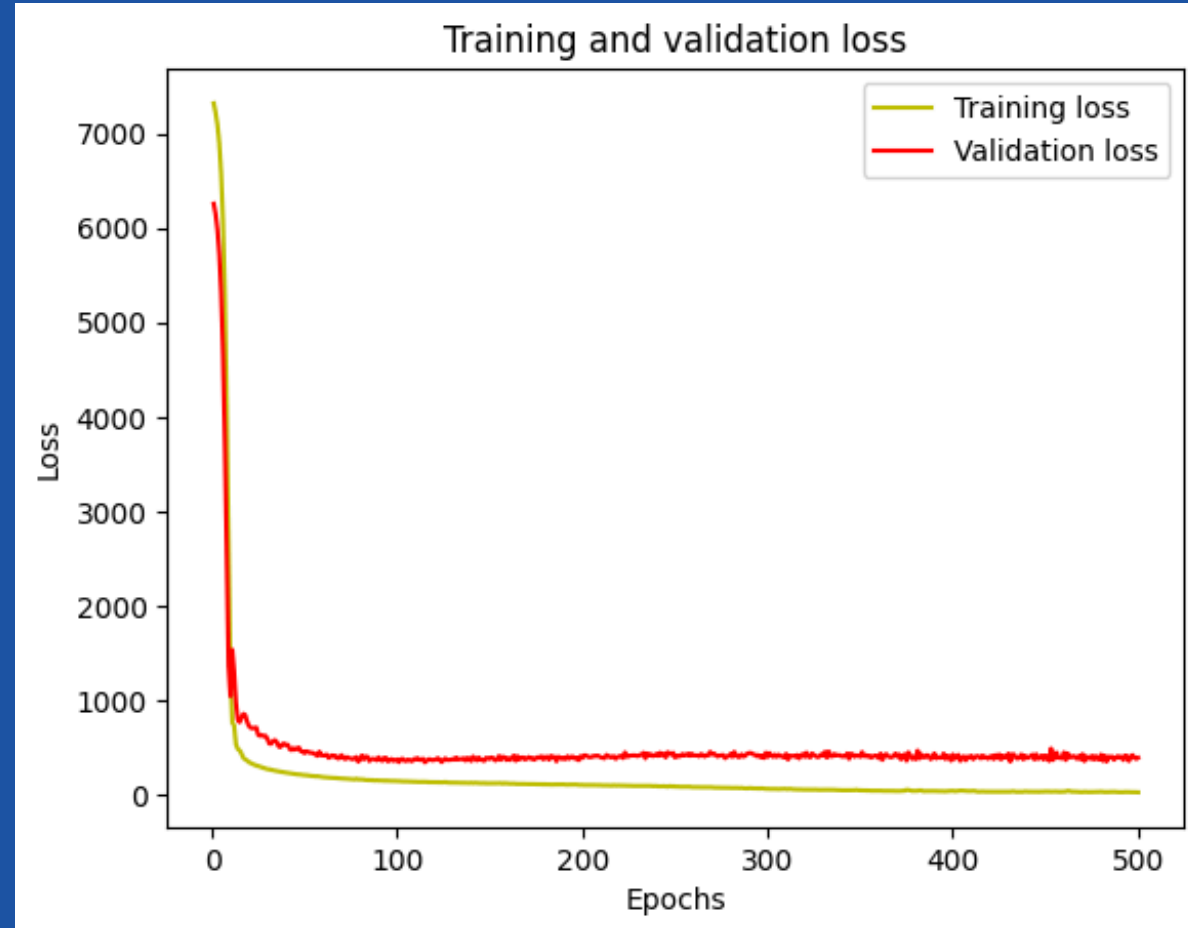
Layer (type)	Output Shape	Param #
=====		
dense_80 (Dense)	(None, 128)	1536
dense_81 (Dense)	(None, 64)	8256
dense_82 (Dense)	(None, 64)	4160
dense_83 (Dense)	(None, 32)	2080
dense_84 (Dense)	(None, 1)	33
=====		
Total params: 16065 (62.75 KB)		
Trainable params: 16065 (62.75 KB)		
Non-trainable params: 0 (0.00 Byte)		

Metodologi

HOMO	float64	0.000000	NA	-7.945000	-4.477000	Column has 5 outliers greater than upper bound (-4.76) or lower than lower bound(-7.51). Cap them or remove them.
LUMO	float64	0.000000	NA	-2614.000000	6.294000	Column has 34 outliers greater than upper bound (-0.42) or lower than lower bound(-3.68). Cap them or remove them.
Gap Energy	float64	0.000000	NA	1.897000	13.536000	Column has 27 outliers greater than upper bound (6.38) or lower than lower bound(1.77). Cap them or remove them.
Dipole Moment	float64	0.000000	NA	0.010000	13.371000	Column has 7 outliers greater than upper bound (9.26) or lower than lower bound(-1.53). Cap them or remove them.
Ionization Potential	float64	0.000000	NA	-4.650000	8.933000	Column has 18 outliers greater than upper bound (7.89) or lower than lower bound(4.39). Cap them or remove them.
Electron Affinity	float64	0.000000	NA	-6.294000	4.655000	Column has 29 outliers greater than upper bound (3.82) or lower than lower bound(0.05). Cap them or remove them., Column has a high correlation with ['Gap Energy']. Consider dropping one of them.
Electronegativity	float64	0.000000	NA	-0.098000	5.971000	Column has 34 outliers greater than upper bound (5.04) or lower than lower bound(3.12). Cap them or remove them., Column has a high correlation with ['Gap Energy', 'Electron Affinity']. Consider dropping one of them.
Global Hardness	float64	0.000000	NA	0.949000	6.768000	Column has 28 outliers greater than upper bound (3.50) or lower than lower bound(0.88). Cap them or remove them., Column has a high correlation with ['Gap Energy', 'Electron Affinity', 'Electronegativity']. Consider dropping one of them.
Global Softness	float64	0.000000	NA	-2.250000	1.054000	Column has 19 outliers greater than upper bound (0.78) or lower than lower bound(0.13). Cap them or remove them.
Electrophilicity	float64	0.000000	NA	-1.212000	19.775674	Column has 2 outliers greater than upper bound (10.25) or lower than lower bound(-6.12). Cap them or remove them.
Fraction of electron transferred	float64	0.000000	NA	-0.831000	1.238000	Column has 7 outliers greater than upper bound (1.21) or lower than lower bound(-0.27). Cap them or remove them.
IE (%)	float64	0.000000	NA	13.000000	98.800000	Column has 27 outliers greater than upper bound (113.97) or lower than lower bound(61.11). Cap them or remove them.

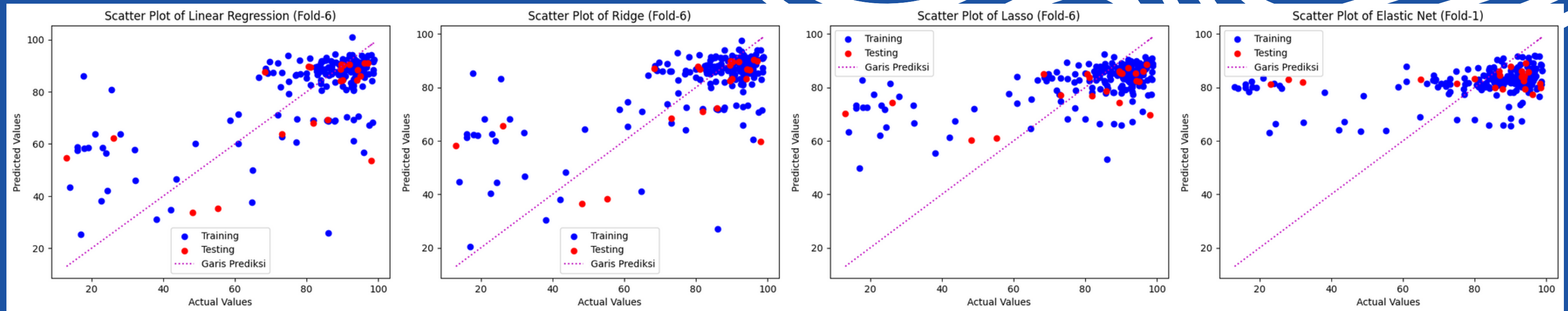


Hasil Penelitian (ANN)



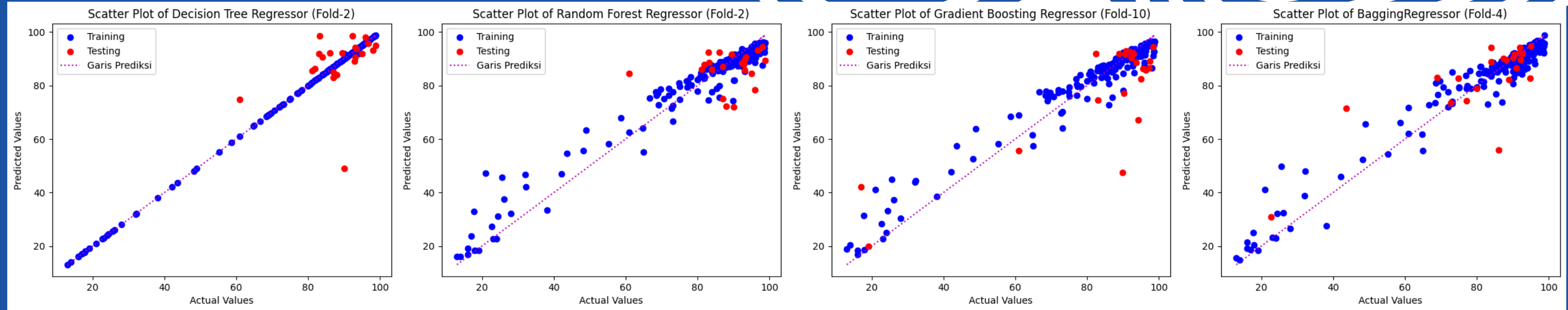
Data	MSE	MAE	R2	RMSE
Pelatihan	33.26	3.50	0.91	5.77
Pengujian	328.41	12.36	0.46	18.12

Hasil Penelitan (Linear)



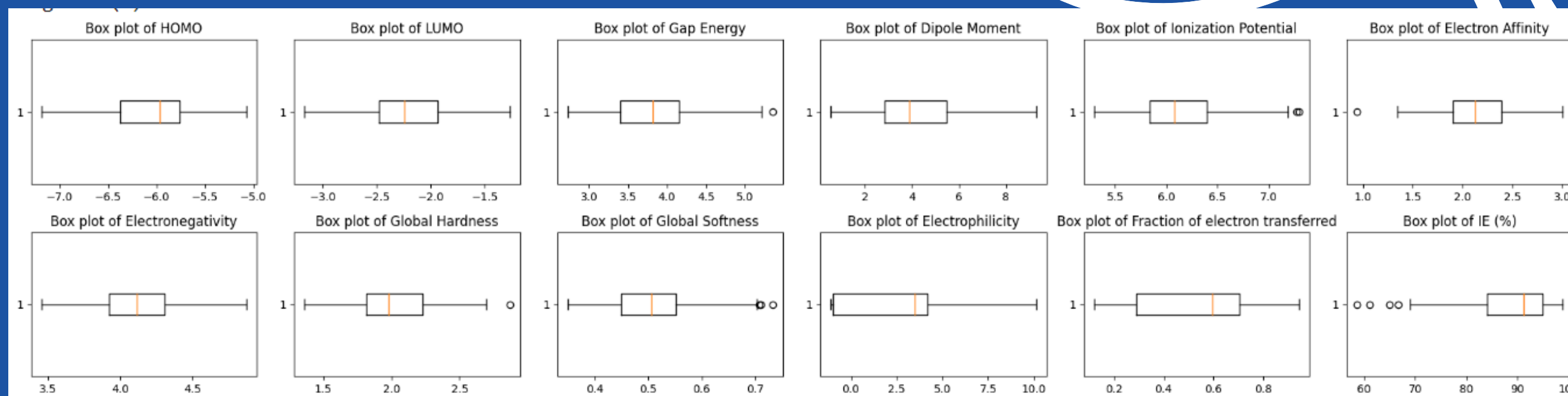
Model	Mean MSE	Mean MAE	Mean RMSE	Mean R ²
Linear	230.35	10.01	15.17	0.48
Ridge	234.95	10.13	15.32	0.47
Lasso	301.86	11.90	17.37	0.32
Elastic Net	385.66	13.34	19.63	0.13

Hasil Penelitian (Non Linear)



Model	Mean MSE	Mean MAE	Mean RMSE	Mean R ²
Decision Tree	367	183	1.817	9.999
Random Forest	23.92	3.25	4.89	9.458
Bagging	31.17	3.46	5.56	9.292
Gradient Boosting	38.93	3.89	6.22	9.117

Hasil Penelitian (Menghapus Outlier)



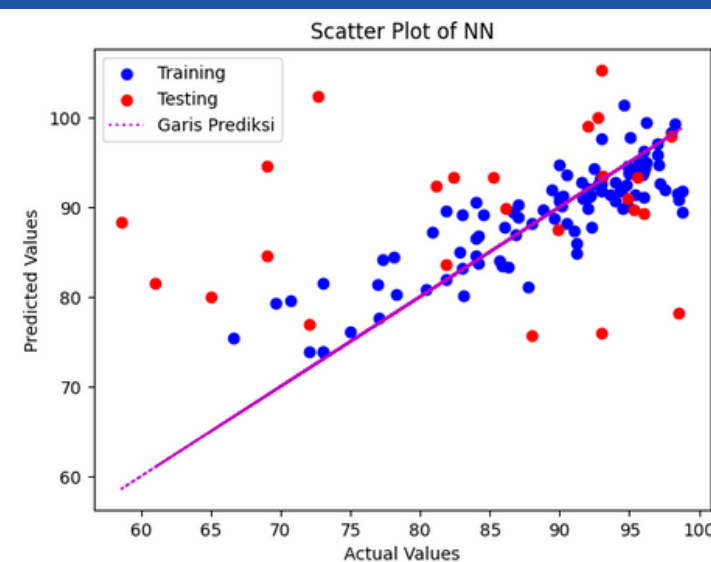
ANN

Train

MSE: 14.462465063356738
MAE: 2.862660321044922
R2: 0.7478144175659773
RMSE: 3.8029547806089856

Test

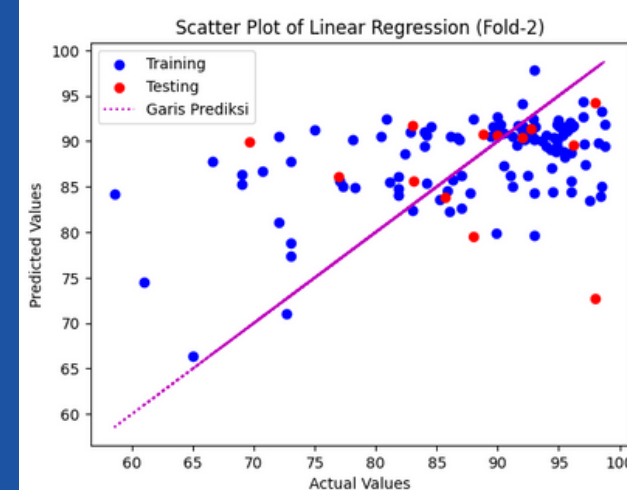
MSE: 194.71742733043249
MAE: 10.975313708496094
R2: -0.33592793008866484
RMSE: 13.954118651152157



MLR

Rata-rata:

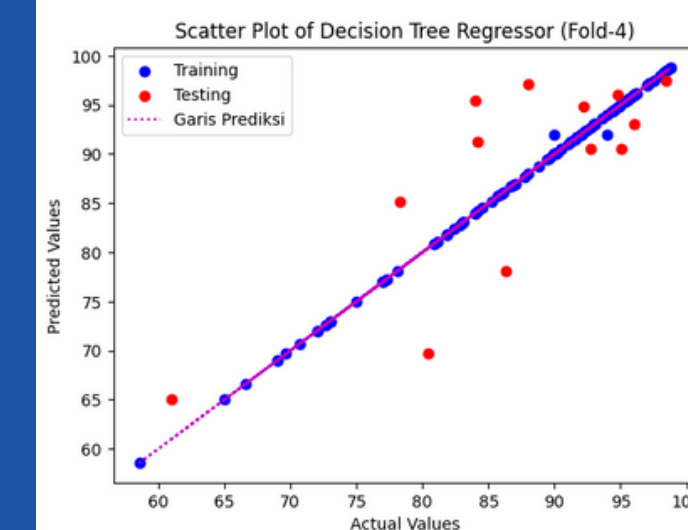
Mean MSE: 60.821407014174234
Mean MAE: 6.0188185722149585
Mean RMSE: 7.797929825042149
Mean R^2: 0.22903050766267535



Decision Tree

Rata-rata:

Mean MSE: 0.05689001264222503
Mean MAE: 0.028445006321112513
Mean RMSE: 0.21333491360300805
Mean R^2: 0.9992775930429909



Kesimpulan

Pada penelitian ini, penulis membandingkan kinerja model machine learning dan deep learning untuk memprediksi efikasi senyawa N-Heterocyclic sebagai corrosion inhibitor. Hasilnya menunjukkan bahwa model non-linear Decision Tree Regressor memiliki nilai akurasi tertinggi dengan r^2 score 0.99 disusul dengan model non-linier lainnya. Sementara itu model ANN masih dibawah non-linier yaitu memiliki r^2 score 0.91. Hal ini mungkin dipengaruhi oleh kompleksitas data, kebutuhan data yang besar untuk ANN, arsitektur NN yang lebih baik, serta kemampuan model non-linear dalam menangani kompleksitas dan mencegah overfitting.