

Contents

Skenario 1 : Default Parameter	2
- Parameter Default :.....	2
- Nilai MSE & RMSE	2
- Data Aktual vs Data Prediksi	3
- Visualisasi Data Aktual & Prediksi	3
- Feature Performance	4
Skenario 2 : GridSearchCV.....	6
- Kajian Teori.....	6
- GridSearchCV Parameter	8
- Performa Parameter	8
- Detail Best Parameters.....	8
- Nilai MSE & RMSE	9

Skenario 1 : Default Parameter

- Parameter Default

bootstrap: True

ccp_alpha: 0.0

criterion: squared_error

max_depth: 10

max_features: 1.0

max_leaf_nodes: None

max_samples: None

min_impurity_decrease: 0.0

min_samples_leaf: 1

min_samples_split: 2

min_weight_fraction_leaf: 0.0

n_estimators: 100

n_jobs: None

oob_score: False

random_state: 42

verbose: 0

warm_start: False

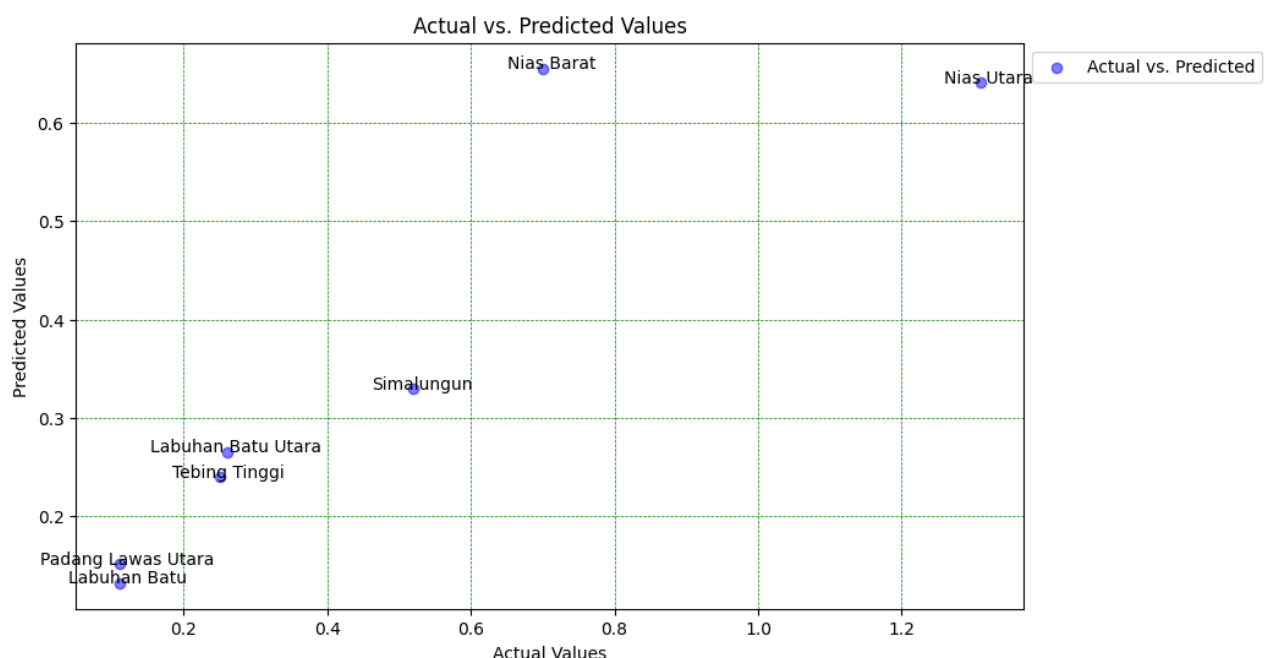
- Nilai MSE & RMSE

Mean Squared Error	Root Mean Squared Error
0.069789	0.264176

- Data Aktual vs Data Prediksi

Indeks Data	Keterangan	Data Aktual (y_actual)	Data Prediksi (y_pred)
31	Tebing Tinggi	0.25	0.23959999999999979
15	Nias Barat	0.7	0.65520000000000003
26	Simalungun	0.52	0.32929999999999987
17	Nias Utara	1.31	0.64080000000000007
8	Labuhan Batu Utara	0.26	0.26549999999999957
9	Labuhan Batu	0.11	0.13199999999999998
19	Padang Lawas Utara	0.11	0.1512

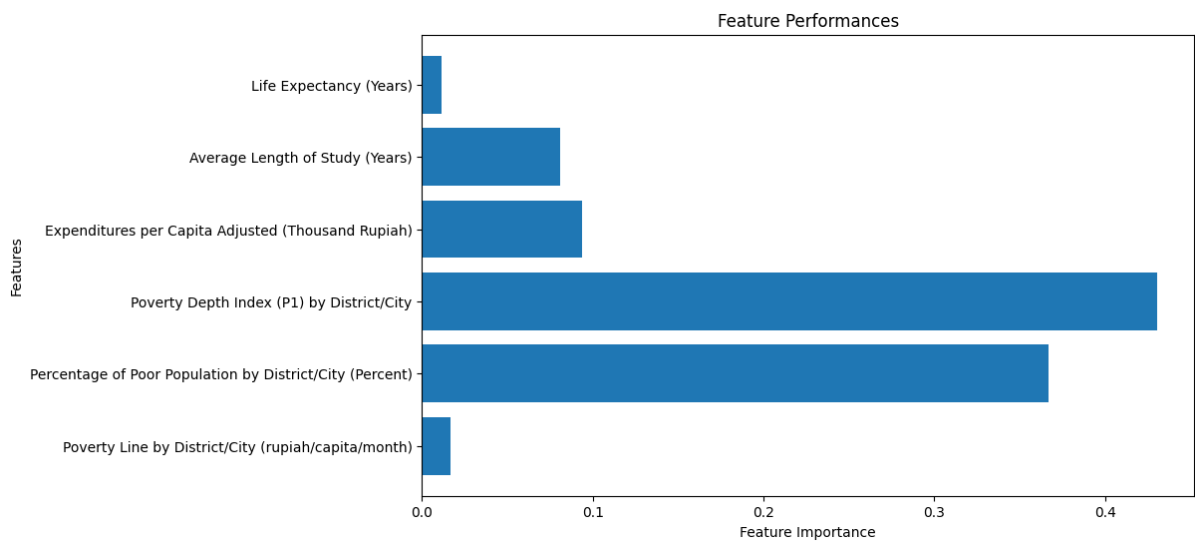
- Visualisasi Data Aktual & Prediksi



Di Kabupaten Tebing Tinggi, model regresi berhasil memberikan prediksi yang cukup mendekati nilai aktual (0.25 vs 0.2396), menunjukkan performa estimasi yang akurat. Temuan serupa terlihat di Kabupaten Nias Barat, dengan perbedaan yang kecil antara nilai aktual dan prediksi (0.7 vs 0.6552), mengindikasikan model yang baik dalam mengevaluasi variabel yang relevan. Namun, Kabupaten Simalungun menunjukkan perbedaan yang lebih besar antara nilai aktual dan prediksi (0.52 vs 0.3293), mengisyaratkan adanya hambatan dalam model dalam memahami hubungan yang kompleks antara variabel. Kabupaten Nias Utara menunjukkan selisih yang signifikan antara nilai aktual dan prediksi (1.31 vs 0.6408), menunjukkan kebutuhan akan perbaikan dalam performa model untuk estimasi yang lebih akurat. Di sisi lain, di Kabupaten Labuhan Batu Utara dan Labuhan Batu, hasil prediksi yang mendekati nilai aktual ditemukan (0.26 vs 0.2655 dan 0.11 vs 0.1320), mencerminkan performa model yang kuat. Terakhir, di Kabupaten Padang Lawas Utara, hasil prediksi (0.1512)

relatif dekat dengan nilai aktual (0.11), namun masih terdapat sedikit perbedaan yang bisa diperbaiki.

- Feature Performance



index	Poverty Line by District/City (rupiah/capita/month)	Percentage of Poor Population by District/City (Percent)	Poverty Depth Index (P1) by District/City	Expenditures per Capita Adjusted (Thousand Rupiah)	Average Length of Study (Years)	Life Expectancy (Years)	Actual	Predicted
31	0.8103498511196232	0.2825366777094179	0.16049382716049385	0.7477275157737139	0.8487544483985767	0.7357142857142867	0.25	0.23959999999999979
15	0.4868530454742497	1.0	0.64444444444444445	0.0	0.19395017793594316	0.569642857142858	0.7	0.6552000000000003
26	0.32438156015264674	0.21959299574065308	0.2962962962962963	0.5744840124050904	0.6672597864768686	0.7794642857142859	0.52	0.32929999999999987
17	0.44088844986746434	0.9361097964978702	1.00000000000000002	0.045235803657362794	0.16014234875444844	0.6116071428571432	1.31	0.6408000000000007
8	0.6305918972121349	0.25887363937529584	0.1679012345679012	0.6529782910918617	0.4911032028469753	0.6124999999999998	0.26	0.26549999999999957
9	0.4940803172324595	0.21959299574065308	0.04938271604938271	0.5691369906961822	0.6263345195729539	0.6535714285714294	0.11	0.13199999999999998
19	0.28600666586126766	0.25177472787505917	0.05925925925925926	0.4593091647952089	0.6370106761565839	0.40000000000000036	0.11	0.1512

Keterangan:

31 = Tebing Tinggi | 15 = Nias Barat | 26 = Simalungun | 17 =Nias Utara | 8 = Labuhan Batu Utara | 9 = Labuhan Batu | 19 = Padang Lawas Utara

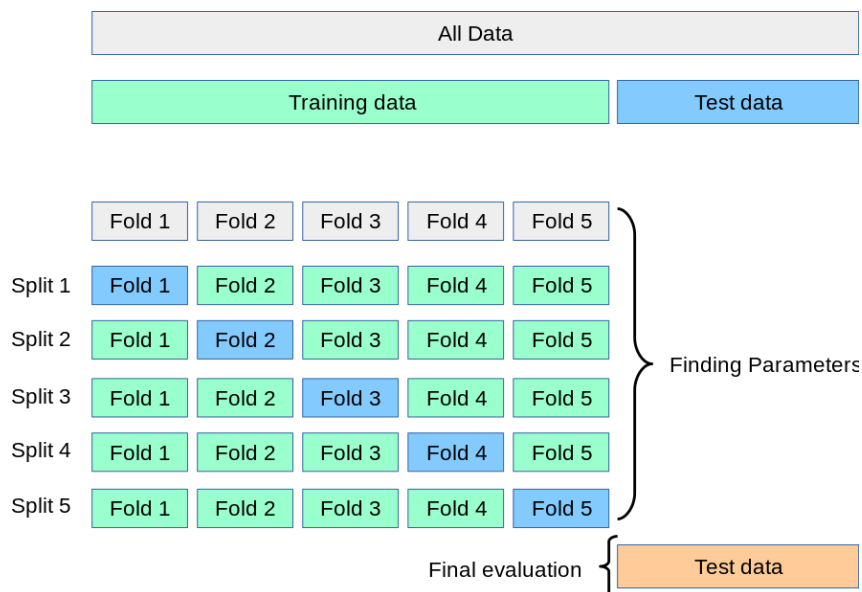
Skenario 2 : GridSearchCV

- Kajian Teori

GridSearchCV adalah metode pembelajaran mesin yang digunakan untuk menemukan kombinasi model dan hyperparameter yang ideal. Tujuan GridSearchCV adalah untuk melakukan tuning hyperparameter, atau mencari nilai hyperparameters yang sesuai untuk meningkatkan kinerja model pembelajaran mesin (Ramadhan et al., 2017). GridSearchCV menggunakan k-fold cross-validation untuk melakukan evaluasi performa model dan hyperparameter tuning secara lebih akurat. GridSearchCV digunakan untuk mencari kombinasi hyperparameter yang optimal, sedangkan k-fold cross-validation digunakan untuk menguji performa model pada data yang tidak terlihat sebelumnya dan menghindari overfitting.

K-fold Cross Validation adalah sebuah metode statistik yang digunakan untuk mengevaluasi kinerja model atau algoritma yang telah dibuat. Proses evaluasi dilakukan dengan membagi dataset menjadi data latih dan data validasi, lalu model akan dilatih menggunakan data latih dan divalidasi dengan data validasi sebanyak *k-fold* kali (Fuadah et al., 2022). Skenario pembagian dataset pada *k-fold cross validation* dapat dilihat pada Gambar 2.3 berikut.

Adapun penjelasan langkah-langkah dalam skenario *k-fold cross validation* adalah



Gambar 2. 1. Skenario *k-fold cross validation*

sebagai berikut :

1. Bagi dataset menjadi *k* subset atau *fold*, biasanya dengan ukuran yang sama.
2. Untuk setiap subset $i=1,2,...,k$:
 - a. Gunakan *subset i* sebagai data uji.
 - b. Gabungkan *subset* lainnya menjadi data latih.

- c. Latih model pada data latih.
 - d. Evaluasi model pada *subset i* dan catat skor evaluasi.
3. Hitung rata-rata skor evaluasi dari *k subset*.

Dalam proses prediksi, metode *k-fold cross validation* dapat digunakan sebagai solusi untuk mengatasi jumlah data yang sedikit. Jumlah data yang tersedia sangat berpengaruh terhadap akurasi algoritma *machine learning*. Jika jumlah data kurang dari 100 *instance*, algoritma *machine learning* mungkin akan memberikan hasil prediksi yang tidak akurat. Untuk meningkatkan akurasi prediksi, algoritma *machine learning* merekomendasikan penggunaan jumlah *instance* yang lebih banyak, yaitu lebih dari 1000 *instance*. (Sena, 2018; Alpaydin, 2014: hlm. 558–559). Adapun nilai *k* yang digunakan dalam sejumlah penelitian ialah *k*=10, namun penggunaan nilai *k*=5 lebih baik karena dapat mengurangi waktu komputasi tanpa mengurangi performa model (Marcot & Hanea, 2021).

Refensi :

Alpaydin, E. (2014). *Introduction to Machine Learning* (3rd ed.). The MIT Press.

Fuadah, Y. N., Ubaidullah, I. D., Ibrahim, N., Taliningsing, F. F., Sy, N. K., & Pramudhito, M. A. (2022). Optimasi Convolutional Neural Network dan K-Fold Cross Validation pada Sistem Klasifikasi Glaukoma. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, 10(3), 728.

Marcot, B. G., & Hanea, A. M. (2021). What is an optimal value of *k* in *k*-fold cross-validation in discrete Bayesian network analysis? *Computational Statistics*, 36(3), 2009–2031.

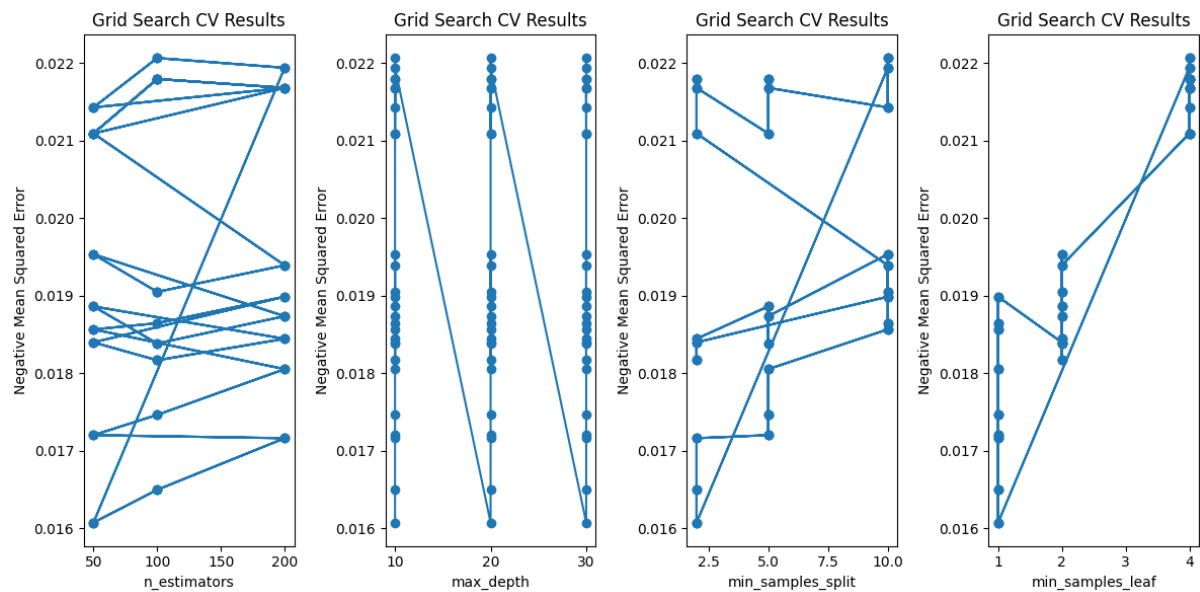
Ramadhan, M. M., Sitanggang, I. S., Nasution, F. R., & Ghifari, A. (2017). Parameter tuning in random forest based on grid search method for gender classification based on voice frequency. *DEStech Transactions on Computer Science and Engineering*, 10(2017).

Sena, S. (2018, March 19). *Pengenalan Deep Learning Part 8 : Gender Classification using Pre-Trained Network (Transfer Learning)*. Medium.Com.

- GridSearchCV Parameter

Parameter	Nilai
n_estimators	50, 100, 200
max_depth	10, 20, 30
min_samples_split	2, 5, 10
min_samples_leaf	1, 2, 4

- Performa Parameter



Berdasarkan Analisa dan visualisasi GridSearch, maka didapatkan best parameters sebagai berikut :

Parameter	Nilai
n_estimators	50
max_depth	10
min_samples_split	2
min_samples_leaf	1

- Detail Best Parameters

bootstrap: True

ccp_alpha: 0.0

criterion: squared_error

max_depth: 10

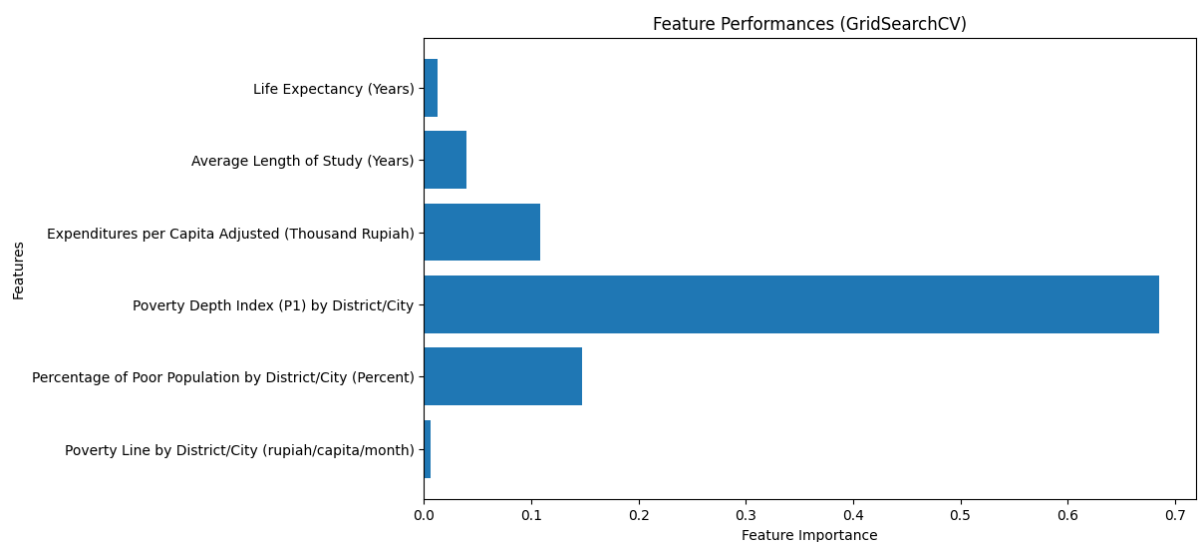
max_features: 1.0

max_leaf_nodes: None
max_samples: None
min_impurity_decrease: 0.0
min_samples_leaf: 1
min_samples_split: 2
min_weight_fraction_leaf: 0.0
n_estimators: 50
n_jobs: None
oob_score: False
random_state: 42
verbose: 0
warm_start: False

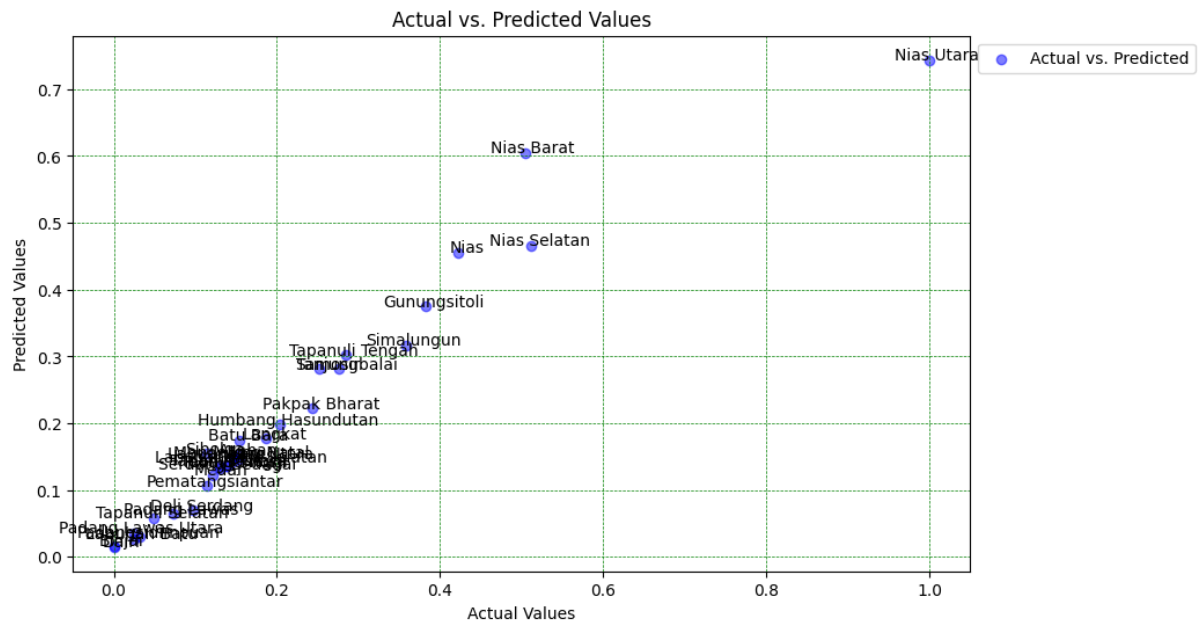
- Nilai MSE & RMSE

Mean Squared Error	Root Mean Squared Error
0.002617	0.156374

- Feature Performance Best Parameter



- Data Aktual vs Data Prediksi (Gridsearch CV)



Keterangan Visualisasi :

Kabupaten: Asahan

Actual Value: 0.15447154471544716

Predicted Value: 0.15284552845528449

Kabupaten: Batu Bara

Actual Value: 0.15447154471544716

Predicted Value: 0.17447154471544707

Kabupaten: Binjai

Actual Value: 0.0

Predicted Value: 0.016910569105691057

Kabupaten: Dairi

Actual Value: 0.0

Predicted Value: 0.014959349593495938

Kabupaten: Deli Serdang

Actual Value: 0.0975609756097561

Predicted Value: 0.07008130081300816

Kabupaten: Gunungsitoli

Actual Value: 0.38211382113821146

Predicted Value: 0.3746341463414636

Kabupaten: Humbang Hasundutan

Actual Value: 0.20325203252032523

Predicted Value: 0.1983739837398373

Kabupaten: Karo

Actual Value: 0.15447154471544716

Predicted Value: 0.14666666666666664

Kabupaten: Labuanbatu Utara

Actual Value: 0.14634146341463417

Predicted Value: 0.14747967479674798

Kabupaten: Labuhan Batu

Actual Value: 0.024390243902439032

Predicted Value: 0.02715447154471546

Kabupaten: Labuhanbatu Selatan

Actual Value: 0.14634146341463417

Predicted Value: 0.14260162601626022

Kabupaten: Langkat

Actual Value: 0.18699186991869918

Predicted Value: 0.17707317073170725

Kabupaten: Mandailing Natal

Actual Value: 0.14634146341463417

Predicted Value: 0.14991869918699194

Kabupaten: Medan

Actual Value: 0.12195121951219512

Predicted Value: 0.12276422764227651

Kabupaten: Nias

Actual Value: 0.4227642276422764

Predicted Value: 0.45593495934959316

Kabupaten: Nias Barat

Actual Value: 0.5040650406504065

Predicted Value: 0.6045528455284552

Kabupaten: Nias Selatan

Actual Value: 0.5121951219512194

Predicted Value: 0.46634146341463384

Kabupaten: Nias Utara

Actual Value: 0.9999999999999999

Predicted Value: 0.7435772357723576

Kabupaten: Padang Lawas

Actual Value: 0.07317073170731708

Predicted Value: 0.06406504065040655

Kabupaten: Padang Lawas Utara

Actual Value: 0.024390243902439032

Predicted Value: 0.035934959349593516

Kabupaten: Padangsidempuan

Actual Value: 0.03252032520325203

Predicted Value: 0.029918699186991856

Kabupaten: Pakpak Bharat

Actual Value: 0.24390243902439027

Predicted Value: 0.2219512195121952

Kabupaten: Pematangsiantar

Actual Value: 0.11382113821138212

Predicted Value: 0.10617886178861793

Kabupaten: Samosir

Actual Value: 0.2520325203252033

Predicted Value: 0.28097560975609753

Kabupaten: Serdang Bedagai

Actual Value: 0.13008130081300812

Predicted Value: 0.13203252032520316

Kabupaten: Sibolga

Actual Value: 0.11382113821138212

Predicted Value: 0.1541463414634147

Kabupaten: Simalungun

Actual Value: 0.3577235772357724

Predicted Value: 0.31674796747967465

Kabupaten: Tanjungbalai

Actual Value: 0.2764227642276422

Predicted Value: 0.2814634146341465

Kabupaten: Tapanuli Selatan

Actual Value: 0.048780487804878064

Predicted Value: 0.058211382113821174

Kabupaten: Tapanuli Tengah

Actual Value: 0.2845528455284553

Predicted Value: 0.3019512195121953

Kabupaten: Tapanuli Utara

Actual Value: 0.13008130081300812

Predicted Value: 0.13642276422764218

Kabupaten: Tebing Tinggi

Actual Value: 0.13821138211382114

Predicted Value: 0.13560975609756112

Kabupaten: Toba Samosir

Actual Value: 0.13821138211382114

Predicted Value: 0.1359349593495936
