

ANALISIS SUPPORT VECTOR REGRESSION
PADA DATA KEMISKINAN 2022

| | |
|---|----|
| A. Analisis Korelasi | 3 |
| B. Linear SVR..... | 4 |
| 1. Parameter Linear SVR | 4 |
| 2. Feature Performances : | 4 |
| 3. NILAI MSE & RMSE : | 5 |
| C. SVR - GridSearch Validation | 5 |
| 1. GridSearchCV Parameter..... | 7 |
| 2. Nilai MSE & RMSE Best Parameters | 7 |
| 3. Performa Parameter | 8 |
| 4. Actual VS Prediction..... | 10 |
| D. RFR VS SVR | 12 |
| 1. Perbandingan nilai MSE & RMSE | 12 |
| 2. Performa Parameter | 12 |
| 3. Actual vs Predicted GridSearch | 12 |

ANALISIS SUPPORT VECTOR REGRESSION
PADA DATA KEMISKINAN 2022

Variabel X (Independen)

| Features |
|--|
| 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) |

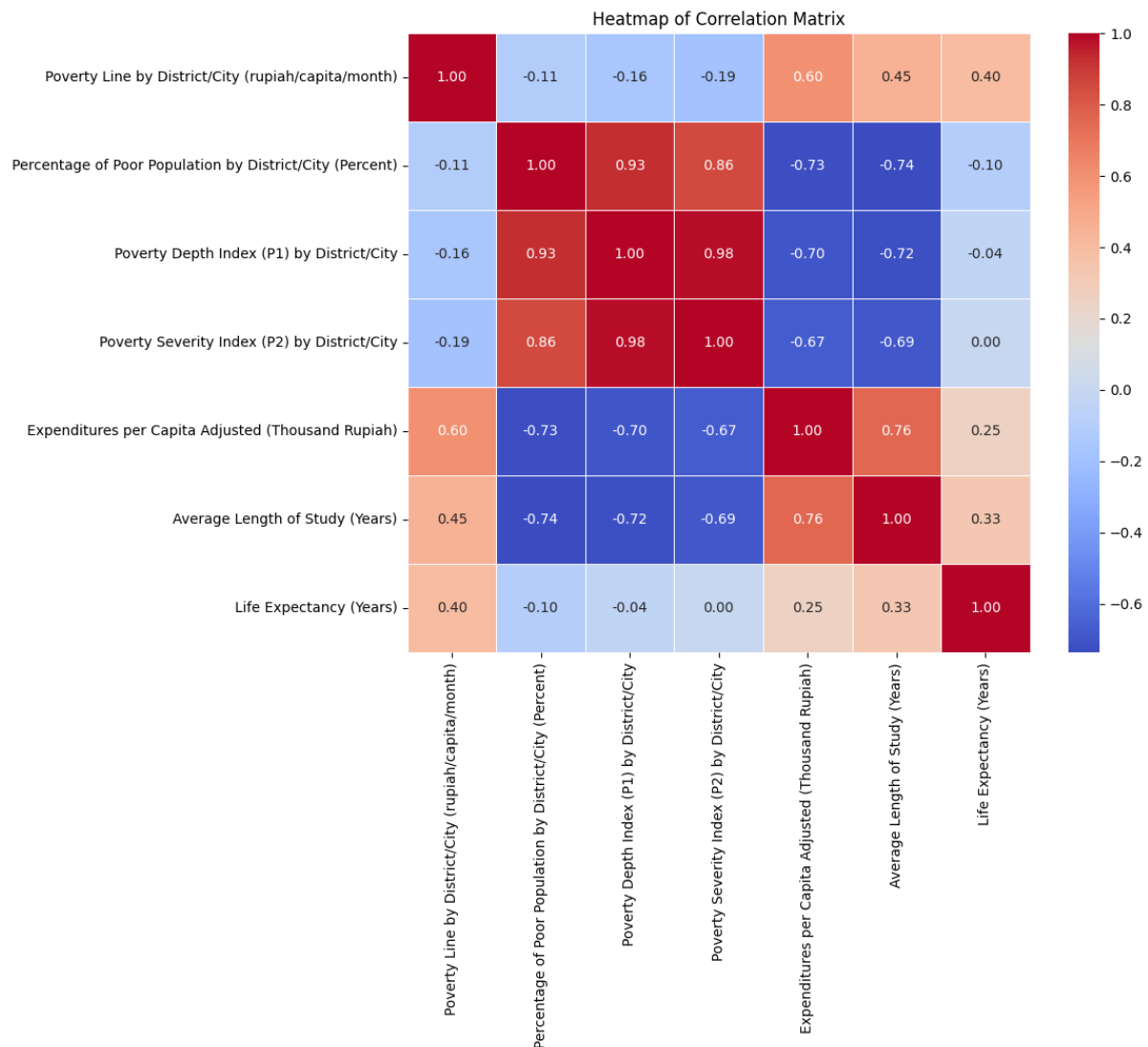
Variabel Y (Dependen)

| Features |
|--|
| Poverty Severity Index (P2) by District/City |

Normalisasi : Min-Max Scaler

$$V' = \frac{v - \min(A)}{\max(A) - \min(A)}$$

A. Analisis Korelasi



Matriks korelasi menyajikan hubungan statistik antara setiap pasangan variabel dalam dataset. Nilai korelasi berkisar dari -1 hingga 1. Korelasi positif, mendekati 1, menandakan hubungan linear positif, di mana kenaikan nilai satu variabel berkorelasi dengan kenaikan nilai variabel lainnya. Sebaliknya, korelasi negatif, mendekati -1, mengindikasikan hubungan linear negatif, di mana peningkatan nilai satu variabel berhubungan dengan penurunan nilai variabel lainnya. Nilai korelasi mendekati 0 menunjukkan kurangnya hubungan linear.

Berdasarkan nilai korelasi diatas, dapat terlihat bahwa beberapa pasangan variabel memiliki korelasi positif yang kuat, sedangkan yang lain memiliki korelasi negatif yang kuat. Hal ini menunjukkan adanya hubungan linear antara beberapa variabel, tetapi juga terdapat hubungan non-linear.

Variabel "Expenditures per Capita Adjusted (Thousand Rupiah)" memiliki korelasi positif yang kuat dengan variabel "Poverty Line by District/City (rupiah/capita/month)" dan "Percentage of Poor Population by District/City (Percent)". Variabel "Percentage of Poor Population by District/City (Percent)" dan "Poverty Depth Index (P1) by District/City" memiliki korelasi positif yang kuat. Variabel "Poverty Severity Index (P2) by District/City" memiliki korelasi

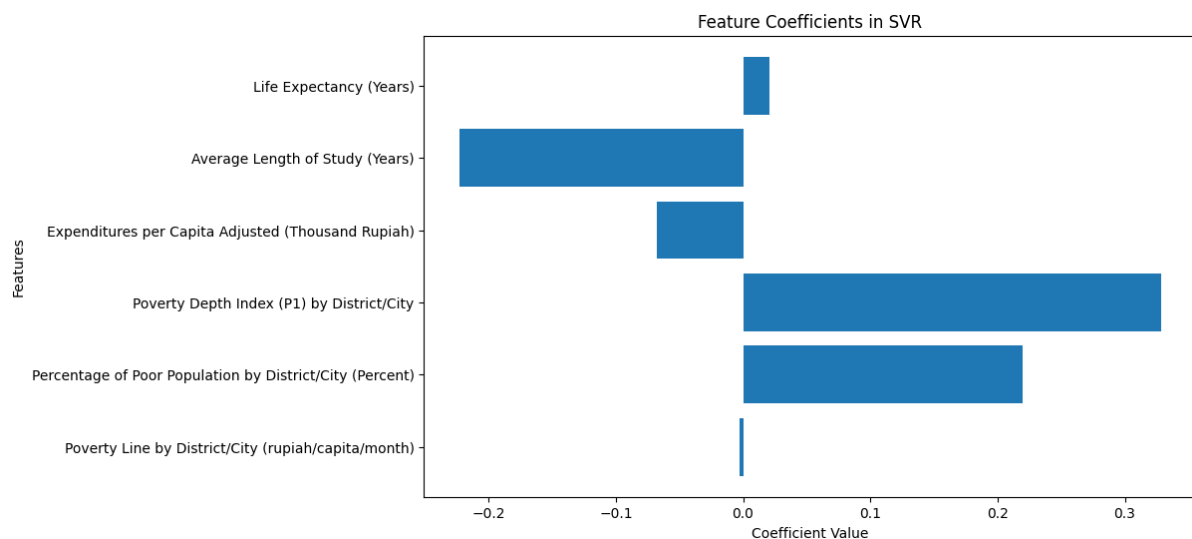
positif yang kuat dengan "Percentage of Poor Population by District/City (Percent)" dan "Poverty Depth Index (P1) by District/City".

B. Linear SVR

1. Parameter Linear SVR:

| Parameter | Nilai |
|------------|--------|
| C | 1 |
| cache size | 200 |
| coef0 | 0 |
| degree | 3 |
| epsilon | 0.1 |
| gamma | scale |
| kernel | linear |
| max_iter | -1 |
| shrinking | true |
| tol | 0.001 |
| verbose | false |

2. Feature Performances :



Dengan menginterpretasikan nilai koefisien yang dihasilkan oleh model SVR, kita dapat mendapatkan pemahaman tentang sejauh mana setiap fitur memengaruhi variabel target (Poverty Severity Index). Berikut adalah interpretasi koefisien untuk setiap fitur:

1. Poverty Line by District/City (rupiah/capita/month): Koefisien -0.00255 menunjukkan bahwa jika Poverty Line meningkat sebesar satu unit, Poverty Severity Index cenderung menurun sebesar 0.00255. Ini menunjukkan adanya hubungan negatif antara Poverty Line dan Poverty Severity Index.

2. Percentage of Poor Population by District/City (Percent): Koefisien 0.21959 menunjukkan bahwa jika Persentase Penduduk Miskin meningkat sebesar satu persen, Poverty Severity Index cenderung meningkat sebesar 0.21959. Ini menunjukkan adanya hubungan positif antara Persentase Penduduk Miskin dan Poverty Severity Index.

3. Poverty Depth Index (P1) by District/City: Koefisien 0.32861 menunjukkan bahwa jika Poverty Depth Index meningkat sebesar satu unit, Poverty Severity Index cenderung meningkat sebesar 0.32861. Ini menunjukkan adanya hubungan positif antara Poverty Depth Index dan Poverty Severity Index.

4. Expenditures per Capita Adjusted (Thousand Rupiah): Koefisien -0.06793 menunjukkan bahwa jika Pengeluaran per Kapita Disesuaikan meningkat sebesar satu ribu Rupiah, Poverty Severity Index cenderung menurun sebesar 0.06793. Ini menunjukkan adanya hubungan negatif antara Pengeluaran per Kapita Disesuaikan dan Poverty Severity Index.

5. Average Length of Study (Years): Koefisien -0.22328 menunjukkan bahwa jika Rata-rata Lama Studi meningkat sebesar satu tahun, Poverty Severity Index cenderung menurun sebesar 0.22328. Ini menunjukkan adanya hubungan negatif antara Rata-rata Lama Studi dan Poverty Severity Index.

6. Life Expectancy (Years): Koefisien 0.02050 menunjukkan bahwa jika Harapan Hidup meningkat sebesar satu tahun, Poverty Severity Index cenderung meningkat sebesar 0.02050. Ini menunjukkan adanya hubungan positif antara Harapan Hidup dan Poverty Severity Index.

3. NILAI MSE & RMSE :

Nilai MSE & RMSE Linear SVR

| Mean Squared Error | Root Mean Squared Error |
|--------------------|-------------------------|
| 0.000605 | 0.024601 |

C. SVR - GridSearch Validation

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 sebagai berikut :

1. Bagi dataset menjadi k subset atau fold, biasanya dengan ukuran yang sama.
2. Untuk setiap subset $i=1,2,\dots,k$:
 - 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).

Referensi :

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.

1. GridSearchCV Parameter

Parameter GridSearch :

| Parameter | Nilai |
|-----------|------------------------------------|
| Kernel | 'linear', 'poly', 'rbf', 'sigmoid' |
| C | 0.1, 1, 10 |
| Epsilon | 0.01, 0.1, 0.2 |

Best Parameter :

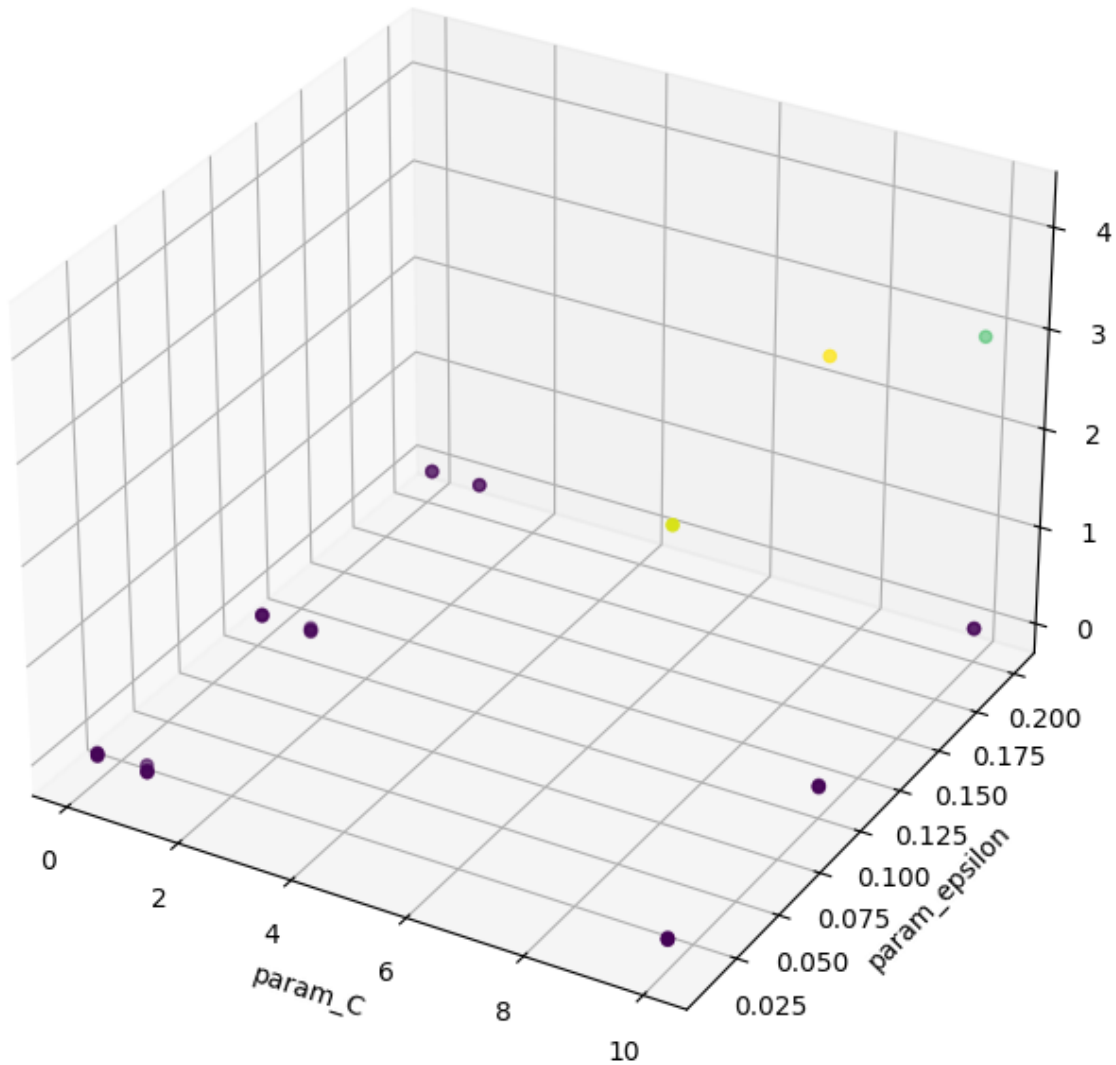
| Parameter | Nilai |
|-----------|----------|
| Kernel | 'linear' |
| C | 10 |
| Epsilon | 0.01 |

2. Nilai MSE & RMSE Best Parameters :

| Mean Squared Error | Root Mean Squared Error |
|--------------------|-------------------------|
| 0.000605 | 0.024601 |

3. Performa Parameter

Grid Search Results for SVR - param_C, param_epsilon, param_kernel



1. **Sumbu X (param1):** Representasi parameter pertama (C) pada sumbu X.
2. **Sumbu Y (param2):** Representasi parameter kedua (ϵ) pada sumbu Y.
3. **Sumbu Z (Mean Test Score):** Representasi skor rata-rata dari uji (Mean Squared Error) pada sumbu Z.

Output nilai :

| param_C | param_epsilon | param_kernel | mean_test_score |
|---------|---------------|--------------|-----------------------|
| 0.1 | 0.01 | linear | 0.016898729826003957 |
| 0.1 | 0.01 | poly | 0.0028744190430186997 |
| 0.1 | 0.01 | rbf | 0.027722620465626367 |
| 0.1 | 0.01 | sigmoid | 0.0388279786949834 |
| 0.1 | 0.1 | linear | 0.023730978165383394 |

| param_C | param_epsilon | param_kernel | mean_test_score |
|---------|---------------|--------------|-----------------------|
| 0.1 | 0.1 | poly | 0.019698756305092137 |
| 0.1 | 0.1 | rbf | 0.030310013015515048 |
| 0.1 | 0.1 | sigmoid | 0.036253446200382365 |
| 0.1 | 0.2 | linear | 0.0364185982756438 |
| 0.1 | 0.2 | poly | 0.03499281327205273 |
| 0.1 | 0.2 | rbf | 0.03865946318221482 |
| 0.1 | 0.2 | sigmoid | 0.03944231804407307 |
| 1 | 0.01 | linear | 0.0026247671578320344 |
| 1 | 0.01 | poly | 0.004150888924311019 |
| 1 | 0.01 | rbf | 0.022374691275294095 |
| 1 | 0.01 | sigmoid | 0.07421715529811987 |
| 1 | 0.1 | linear | 0.01189853731380901 |
| 1 | 0.1 | poly | 0.019698756305092137 |
| 1 | 0.1 | rbf | 0.031064851397126426 |
| 1 | 0.1 | sigmoid | 0.05137190166627535 |
| 1 | 0.2 | linear | 0.03829012976629646 |
| 1 | 0.2 | poly | 0.03499281327205273 |
| 1 | 0.2 | rbf | 0.0435447650586138 |
| 1 | 0.2 | sigmoid | 0.0380267958208544 |
| 10 | 0.01 | linear | 0.0010693144464942882 |
| 10 | 0.01 | poly | 0.008740407690324088 |
| 10 | 0.01 | rbf | 0.022436209058709165 |
| 10 | 0.01 | sigmoid | 3.971117668507837 |
| 10 | 0.1 | linear | 0.008053774647147672 |
| 10 | 0.1 | poly | 0.019698756305092137 |
| 10 | 0.1 | rbf | 0.031064851397126426 |
| 10 | 0.1 | sigmoid | 4.228620866835344 |
| 10 | 0.2 | linear | 0.03829012976629646 |
| 10 | 0.2 | poly | 0.03499281327205273 |
| 10 | 0.2 | rbf | 0.0435447650586138 |
| 10 | 0.2 | sigmoid | 3.0068478632659668 |

Hasil Grid Search untuk parameter SVR menunjukkan variasi performa model dengan kombinasi berbagai nilai parameter C, epsilon, dan jenis kernel. Pada hasil ini, kita dapat mengidentifikasi pola tertentu yang memberikan wawasan tentang pengaruh parameter terhadap kinerja model SVR pada dataset tertentu.

Pengaruh Parameter C:

- Pada nilai C sebesar 0.1, kita melihat variasi hasil yang signifikan tergantung pada jenis kernel yang digunakan. Kernel linear dan poly memberikan MSE yang lebih rendah dibandingkan dengan kernel rbf dan sigmoid. Artinya, pada nilai C yang rendah, kompleksitas model lebih rendah, dan model lebih cenderung memberikan hasil yang baik dengan jenis kernel linear dan poly.
- Saat C meningkat menjadi 1 dan 10, kita melihat tren penurunan MSE untuk kernel linear dan poly. Ini menunjukkan bahwa meningkatkan kompleksitas model dengan nilai C yang lebih tinggi dapat meningkatkan kinerja model, terutama dengan penggunaan kernel linear dan poly.

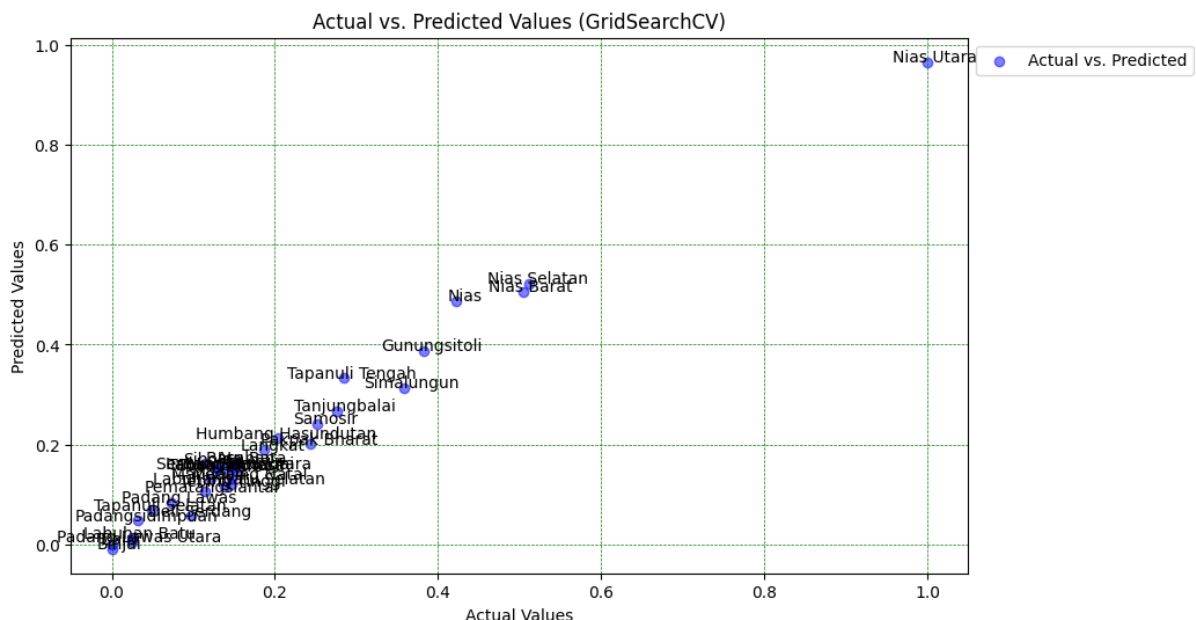
Pengaruh Parameter Epsilon:

- Tidak terdapat pola yang konsisten dalam pengaruh nilai parameter **epsilon** terhadap MSE. Secara teoritis, **epsilon** mengontrol toleransi model terhadap kesalahan prediksi. Dalam hasil ini, variasi nilai epsilon tidak secara jelas mempengaruhi performa model, menunjukkan kompleksitas model yang serupa pada rentang epsilon yang diuji..

Pengaruh Parameter Kernel:

- Kernel linear dan poly cenderung memberikan MSE yang lebih rendah dibandingkan dengan kernel rbf dan sigmoid pada nilai C yang rendah dan sedang. Ini mencerminkan kecenderungan bahwa model SVR dengan kernel linear dan poly cocok untuk dataset ini.

4. Actual VS Prediction



| kabupaten | actual | predicted |
|-----------|---------------------|---------------------|
| Asahan | 0.15447154471544716 | 0.16405019356360734 |
| Batu Bara | 0.15447154471544716 | 0.16408079259235636 |

| | | |
|------------------------|----------------------|------------------------|
| Binjai | 0 | -0.00954659974766698 |
| Dairi | 0 | -0.0002225783643830595 |
| Deli Serdang | 0.0975609756097561 | 0.057265088263891366 |
| Gunungsitoli | 0.38211382113821146 | 0.38812735350920663 |
| Humbang Hasundutan | 0.20325203252032523 | 0.2128771434496936 |
| Karo | 0.15447154471544716 | 0.1459336018439909 |
| Labuanbatu Utara | 0.14634146341463417 | 0.15207986729106254 |
| Labuhan Batu | 0.024390243902439032 | 0.01236492683336892 |
| Labuhanbatu Selatan | 0.14634146341463417 | 0.12091660401466336 |
| Langkat | 0.18699186991869918 | 0.18976030916176073 |
| Mandailing Natal | 0.14634146341463417 | 0.13176875749198297 |
| Medan | 0.12195121951219512 | 0.13054638079720776 |
| Nias | 0.4227642276422764 | 0.48798635062858325 |
| Nias Barat | 0.5040650406504065 | 0.5059775807036678 |
| Nias Selatan | 0.5121951219512194 | 0.5226486103034212 |
| Nias Utara | 0.9999999999999999 | 0.9662437816825288 |
| Padang Lawas | 0.07317073170731708 | 0.08356283464266306 |
| Padang Lawas Utara | 0.024390243902439032 | 0.00476772577485101 |
| Padangsidimpuan | 0.03252032520325203 | 0.04755238032427827 |
| Pakpak Bharat | 0.24390243902439027 | 0.2014682192994321 |
| Pematangsiantar | 0.11382113821138212 | 0.10542188017705983 |
| Samosir | 0.2520325203252033 | 0.2417313151013864 |
| Serdang Bedagai | 0.13008130081300812 | 0.15131039785996764 |
| Sibolga | 0.11382113821138212 | 0.16131197301472783 |
| Simalungun | 0.3577235772357724 | 0.3131895566036124 |
| Tanjungbalai | 0.2764227642276422 | 0.26661241251247064 |
| Tapanuli Selatan | 0.048780487804878064 | 0.06789714585759826 |
| Tapanuli Tengah | 0.2845528455284553 | 0.33342123734495205 |
| Tapanuli Utara | 0.13008130081300812 | 0.14998110302069806 |
| Tebing Tinggi | 0.13821138211382114 | 0.11613252009447267 |
| Toba Samosir | 0.13821138211382114 | 0.14953966719061812 |

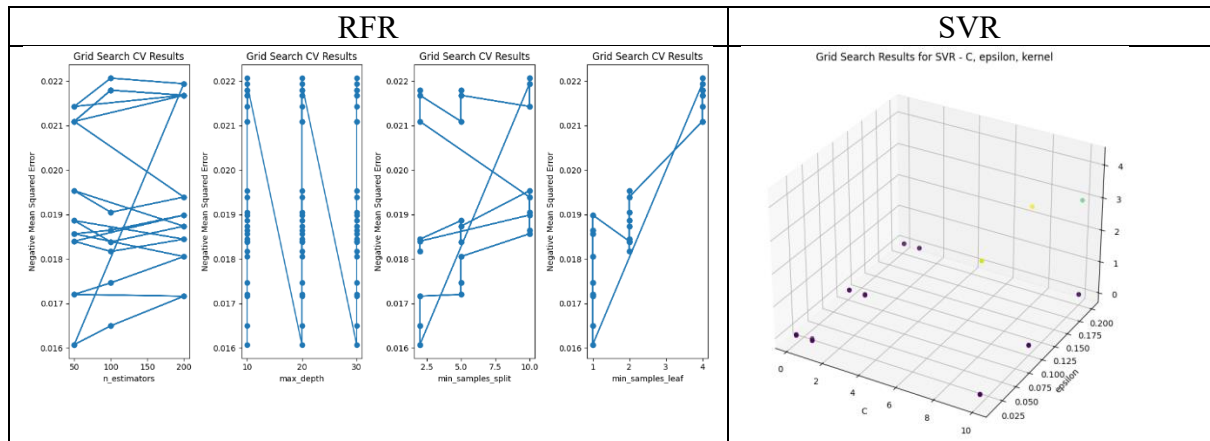
D. RFR VS SVR

1. Perbandingan nilai MSE & RMSE

Nilai MSE & RMSE antara RFR dan SVR menggunakan metode GridSearch Validation

| RFR | | SVR | |
|----------|----------|----------|----------|
| MSE | RMSE | MSE | RMSE |
| 0.002617 | 0.156374 | 0.000605 | 0.024601 |

2. Performa Parameter



3. Actual vs Predicted GridSearch

