

NASKAH PUBLIKASI (*MANUSCRIPT*)

**KOMPARASI EVALUASI PARAMETER KERNEL LINIER
SUPPORT VECTOR REGRESSION (SVR) UNTUK
AUTOMATED ESSAY SCORING (AES) BAHASA
INDONESIA**

***COMPARATION EVALUATION OF LINIER SUPPORT
VECTOR REGRESSION (SVR) PARAMETERS FOR
INDONESIAN LANGUAGE AUTOMATED ESSAY SCORING
(AES)***

Riski Nopiansyah, Naufal Azmi Verdhika



DISUSUN OLEH :

RISKI NOPIANSYAH

1811102441083

**PROGRAM STUDI S1 TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR
SAMARINDA**

2022

Naskah Publikasi (*Manuscript*)

**Komparasi Evaluasi Parameter Kernel Linier Support
Vector Regression (SVR) untuk Automated Essay Scoring
(AES) Bahasa Indonesia**

***Comparison Evaluation of Linier Support Vector
Regression (SVR) Parameters for Indonesian Language
Automated Essay Scoring (AES)***

Riski Nopiansyah, Naufal Azmi Verdhika



Disusun Oleh :

Riski Nopiansyah

1811102441083

**PROGRAM STUDI S1 TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR
SAMARINDA**

2022

HALAMAN PENGESAHAN

KOMPARASI EVALUASI PARAMETER *KERNEL LINEAR SUPPORT VECTOR
REGRESSION (SVR)* UNTUK *AUTOMATED ESSAY SCORING (AES)* BAHASA
INDONESIA

NASKAH PUBLIKASI

DISUSUN OLEH :

RISKI NOPIANSYAH

1811102441083

Dosen Pembimbing



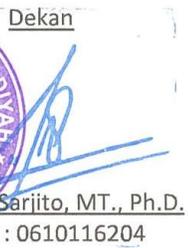
Naufal Azmi Verdikha, S.Kom., M. Eng
NIDN : 1114048801

Penguji



Asslia Johar Latipah, M.Cs
NIDN : 1124098902

Dekan



Prof. N. Sarjito, MT., Ph.D.
NIDN : 0610116204

Ketua Progam Studi



Asslia Johar Latipah, M.Cs
NIDN : 1124098902

**Komparasi Evaluasi Parameter Kernel Linier Support Vector Regression (SVR)
untuk Automated Essay Scoring (AES) Bahasa Indonesia**

Riski Nopiansyah¹, Naufal Azmi Verdhika²

^{1,2}Teknik Informatika, Universitas Muhammadiyah Kalimantan Timur, Jl. Ir. H. Juanda
No.15, Sidodadi, Kec.Samarinda Ulu, Kota Samarinda, Kalimantan Timur 75124,
Indonesia

**Email Corresponding Author: riski.rz1996@gmail.com*

ABSTRAK

Automated Essay Scoring (AES) merupakan sebuah sistem yang dapat memberikan sebuah penilaian terhadap soal ujian essay Bahasa Indonesia maupun Bahasa Asing secara otomatis dengan membandingkan kunci jawaban dengan jawaban yang diberikan oleh pelajar. Natural Language Processing (NLP) adalah salah satu bagian dari bahasa alami yang merupakan cabang dari ilmu komputer yang membahas tentang interaksi antara manusia dengan komputer menggunakan bahasa manusia. Penelitian ini bertujuan mencari nilai evaluasi Root Mean Squared Error (RMSE) menggunakan ekstraksi fitur Term Frequency - Inverse Document Frequency (TF-IDF) dan di regresi menggunakan Support Vector Regression (SVR) dengan kernel linier. Hasil evaluasi tersebut kemudian dibandingkan dengan hasil evaluasi pada penelitian terdahulu yang menggunakan SVR dengan kernel RBF (Radial Basis Function) dengan hasil 2,166, Hasil evaluasi RMSE pada penelitian ini sebanyak 2,151 sehingga dapat disimpulkan bahwa penggunaan SVR dengan kernel linier menghasilkan nilai evaluasi lebih baik dibandingkan penelitian sebelumnya.

Kata kunci: *Essay Scoring, Bahasa Indonesia, Natural Language Processing, Parameter kernel linier, Support Vector Regression.*

Comparison Evaluation of Linier Support Vector Regression (Svr) Parameters for Indonesian Language Automated Essay Scoring (Aes)

Riski Nopiansyah¹, Naufal Azmi Verdhika²

^{1,2}Teknik Informatika, Universitas Muhammadiyah Kalimantan Timur, Jl. Ir. H. Juanda No.15, Sidodadi, Kec.Samarinda Ulu, Kota Samarinda, Kalimantan Timur 75124, Indonesia

**Email Corresponding Author: riski.rz1996@gmail.com*

ABSTRACT

Automated Essay Scoring (AES) is a system that can provide an assessment of Indonesian and foreign language essay exam questions automatically by comparing the answer keys with the answers given by students. Natural Language Processing (NLP) is a part of natural language which is a branch of computer science that discusses the interaction between humans and computers using human language. This study aims to find the evaluation value of Root Mean Squared Error (RMSE) using Term Frequency - Inverse Document Frequency (TF-IDF) feature extraction and regression using Support Vector Regression (SVR) with a linear kernel. The evaluation results were then compared with the evaluation results in previous studies using SVR with RBF (Radial Basis Function) kernel with a result of 2.166. previously.

Keywords: *Automated Essay Scoring, Indonesia Language, liniear kernel Parameter, Natural Language Processing, Support Vector Regression.*

1. Pendahuluan

Natural Language Processing (NLP) adalah bagian dari kecerdasan buatan yang berfokus pada interpretasi komputer dalam bahasa Indonesia yang banyak diterapkan mulai dari sekolah hingga perguruan tinggi yang sedang melaksanakan pembelajaran online. Contoh dari penggunaan NLP yang telah diterapkan oleh dunia pendidikan adalah *Automated Essay Scoring* (AES) (Ramesh & Sanampudi, 2021).

Automated Essay Scoring (AES) sendiri telah digunakan oleh beberapa peneliti, salah satunya pada tahun 2019 oleh (Maulidya, 2019), menerapkan sistem *Automated Essay Scoring* Berbasis Web, tujuannya untuk mengevaluasi sistem pembelajaran secara daring, adapun metode yang telah digunakan yaitu metode *Latent Semantic Analysis* (LSA) (Maulidya, 2019), fungsi dari LSA sendiri disini untuk mengekstrak dan mempresentasikan suatu kalimat dengan perhitungan secara matematis.

pada tahun 2021 (Thamrin dkk) melakukan penelitian *Automated Essay Scoring* (AES) di Universitas Muhammadiyah Kalimantan Timur yang berjudul "*Text Classification and Similarity Algorithms in Essay Grading*". Dengan tujuan untuk melakukan penilaian essay dengan teknik klasifikasi dan algoritma kesamaan menggunakan *Support Vector Machine* (SVM) dengan *Term Frequency - Inverse Document Frequency* (TF-IDF) dan LSA, hasil dari penelitiannya menggunakan SVM dan TF-IDF akan dihitung menggunakan *Root Mean Square Error* (RMSE).

Pada tahun 2021 (Verdikha dkk) melakukan penelitian yang berjudul "*Regression and Oversampling Method for Indonesian Language Automated Essay Scoring*". Dengan tujuan mengurangi resiko kesalahan model pembelajaran yang terdapat pada penelitian Thamrin sebelumnya. Dalam penelitiannya, Verdikha menggunakan tiga metode *regresi*, yaitu *Support Vector Regression* (SVR) dengan *kernel* RBF, *Logistic Regression* (LR), *MLP-Regression* (MLP-R) dan mencoba untuk menggunakan *oversample dataset* menggunakan (SMOTE) agar dapat menyeimbangkan data-data penelitiannya.

Dalam penelitian (Verdikha dkk., 2021) untuk melakukan perbandingan sama seperti yang dilakukan oleh Thamrin, yaitu menggunakan nilai evaluasi *Root Mean Square Error* (RMSE), yang mana nilai yang mendekati NOL adalah nilai yang terbaik. Dalam penelitian Verdikha dengan menggunakan metode SVR nilai yang dihasilkan RMSE lebih baik dibandingkan penelitian (Thamrin dkk., 2021) yaitu 2,16. Hasil ini menunjukkan bahwa metode SVR menggunakan *kernel* RBF lebih baik dari pada klasifikasi metode untuk model pembelajaran yang sudah diteliti sejauh ini.

Metode SVR memiliki beberapa parameter *kernel*, yaitu *kernel* Polynomial, *kernel* Sigmoid, *kernel* Radial Basis Function (RBF), dan *Kernel linier*, pada penelitian ini peneliti memilih untuk menggunakan metode SVR dengan *Kernel linier* karena pada penelitian sebelumnya yang dilakukan oleh (Pangarkar dkk., 2020) terbukti bahwa dengan menggunakan *kernel linier* mendapatkan nilai RMSE terbaik dibandingkan *kernel-kernel* lainnya.

Sehingga pada penelitian ini penulis akan melakukan komparasi untuk membandingkan nilai dari penelitian sebelumnya yang dilakukan oleh Verdikha dengan menggunakan metode ekstraksi fitur TF-IDF dan metode SVR dengan parameter *kernel linier*. Dalam mengukur tingkat kesalahan hasil dari penelitian di sini menggunakan nilai evaluasi yang sama dengan penelitian Verdikha, yaitu *Root Mean Square Error* (RMSE).

2. Metode Penelitian

2.1 *Term Frequency-Inverse Document Frequency* (TF-IDF)

TF-IDF merupakan pembobotan yang digunakan dalam pengambilan suatu informasi dalam *text mining*. Bobot TF-IDF digunakan untuk mengevaluasi seberapa pentingnya sebuah kata di dalam sekumpulan data. Adapun pembobotan TF-IDF sendiri terdiri dari dua faktor, yaitu:

1. *Term Frequency* (TF)

Term Frequency (TF) adalah suatu kemunculan f_l di dalam sebuah dokumen d_j yang di bandingkan dengan f_l yang sering muncul pada dokumen itu (Dalimunthe & Hayadi, 2022).

2. *Inverse Document Frequency* (IDF)

Inverse Document Frequency (IDF) adalah *Frequency* kemunculan suatu istilah f_i di dalam seluruh dokumen. Jika jumlah keseluruhan dokumen pada sebuah data dinyatakan dengan nilai N dan jumlah dokumen yang memiliki istilah f_i dinyatakan dengan n_i , maka nilai IDF _{i} -nya dapat dinyatakan dengan:

$$idf(t) = \log \frac{1+n}{1+df(t)} + 1 \tag{1}$$

Keterangan:

idf = *invers document frekuensi*

t = *term* (kata)

n = *total document*

Untuk menentukan nilai antara TF dan IDF diperlukan penggabungan perhitungan antara TF dengan IDF, contohnya sebagai berikut:

$$tf - idf(t, d) = tf(t, d) \times idf(t) \tag{2}$$

Keterangan:

tf = *term Frequency*

idf = *inverse document frequency*

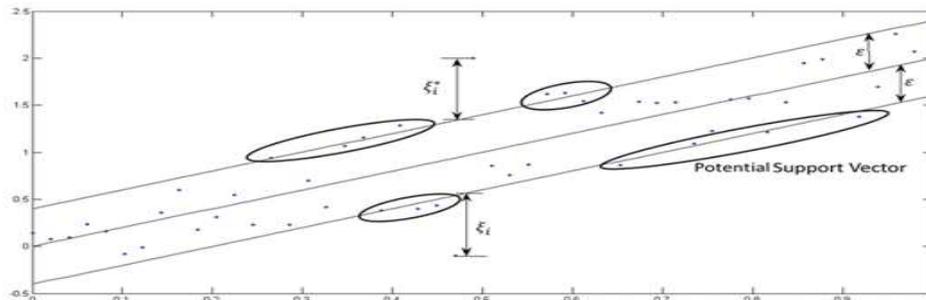
t = *term*

d = *document*

Hasil dari perhitungan TF IDF selanjutnya akan dilakukan tahapan normalisasi, Menurut (Maulana dkk., 2019), Normalisasi memiliki tujuan untuk mendapatkan data dengan ukuran yang lebih kecil yang dapat mewakili data yang asli tanpa kehilangan karakteristiknya sendiri.

2.2 *Support Vector Regression* (SVR)

Support Vector Regression (SVR) adalah penerapan dari SVM yang digunakan untuk kasus *regresi* yang mana outputnya berupa bilangan riil atau kontinu. Ide dasar dari SVR sendiri adalah untuk menentukan *dataset* yang mana datanya akan di bagi menjadi *data training* dan *data testing*. Kemudian dari dua data tersebut akan dilakukan suatu fungsi *regresi* dengan Batasan definisi tertentu sehingga dapat memiliki prediksi yang memiliki nilai aktual (Lestari dkk., 2021). Pada metode SVM, penerapannya di kasus klasifikasi dengan menghasilkan nilai bulat, sedangkan pada metode SVR, penerapannya di kasus *regresi* yang menghasilkan bilangan riil.



Gambar 1 Ilustrasi SVR

Pada Gambar 1 di atas menjelaskan bahwa, garis tengah yang biasa disebut *hyperplane* diapit dengan dua garis pembatas yang memiliki nilai “- & +” seperti gambar diatas, adapun simbol ϵ di atas sebagai jarak antar *hyperplane* dengan pembatas, Titik-titik yang dilingkari adalah *potential support vectors* dan titik titik yang berada pada garis *hyperplane* dan garis pembatas maupun yang berada diluar garis pembatas adalah *data points* yang dapat menjadi calon pembatas, sehingga keseluruhan *data points* dapat masuk menjadi satu klaster, sehingga tetap dapat meminimalisir nilai ϵ nya. Jika divisualisasikan, garis *hyperplane* sebisa mungkin melewati semua titik-titik *data points* tersebut.

2.3 Kernel

Dalam penelitian (Asyiva, 2019) Fungsi dari *kernel* sangat penting dalam penggunaan metode SVR, *Kernel* sendiri sangat membantu mengatasi permasalahan *non-linear* pada dimensi tinggi yang harus dilakukan yaitu mengganti inner product (x_i dan x_j) dengan fungsi *kernel* yang telah dipilih, Karena kinerja dari metode SVR ditentukan oleh jenis fungsi *kernel* dan parameter yang akan digunakan, adapun *kernel* yang digunakan dalam penelitian ini adalah:

1) Kernel linier

Kernel linier merupakan fungsi *kernel* yang paling sederhana. digunakan ketika data yang dianalisis sudah terpisah secara *linear*. *kernel Linear* cocok ketika terdapat banyak fitur dikarenakan pemetaan ke ruang dimensi yang lebih tinggi tidak benar – benar meningkatkan kinerja seperti pada klasifikasi teks (Julianto dkk., 2022).

$$K(x, y) = x \cdot y \quad (3)$$

Keterangan:

$K(x, y)$ = nilai *kernel* dari data x dan data y
 x = nilai data 1
 y = nilai data 2

2) Kernel RBF

Kernel RBF merupakan fungsi *kernel* yang biasa digunakan dalam analisis ketika data tidak terpisah secara *linear*. RBF *kernel* memiliki dua parameter yaitu *Gamma* dan *Cost*. Parameter *Cost* atau biasa disebut sebagai C merupakan parameter yang bekerja sebagai pengoptimalan SVM untuk menghindari misklasifikasi di setiap sampel dalam training *dataset*. Parameter *Gamma* menentukan seberapa jauh pengaruh dari satu sampel training *dataset* dengan nilai rendah berarti “jauh”, dan nilai tinggi berarti “dekat”(Julianto dkk., 2022). Fungsi *kernel* RBF menghitung antara dua vektor, rumus persamaan yang digunakan menggunakan rumus persamaan 4 sebagai berikut:

$$k(x, y) = \exp(-\gamma \|x - y\|^2) \quad (4)$$

Dimana x dan y adalah input vektor. γ dikenal sebagai kemiringan. *Kernel* RBF dikenal juga *kernel* varian *Gaussian* (Barupal & Fiehn, 2019).

2.4 Root Mean Square Error (RMSE)

RMSE digunakan untuk menghitung keakuratan nilai yang direkomendasikan oleh sistem. Semakin banyak nilai rekomendasi sistem yang sama dengan nilai sebenarnya (penilaian oleh pengajar) menunjukkan semakin akuratnya hasil dari sistem. Semakin kecil nilai yang dihasilkan maka semakin bagus pula hasil peramalan yang dilakukan (Cholis dkk., 2019). Adapun rumusan dari RMSE sebagai berikut:

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n}} \quad (5)$$

Keterangan:

- \hat{y}_i = Nilai hasil peramalan
- y_i = Nilai aktual / Nilai sebenarnya
- n = Jumlah data
- l = Urutan data pada dokumen

2.5 Data

Data yang digunakan dalam penelitian AES diambil dari hasil jawaban esai pelajaran Bahasa Indonesia pada semester 2 tahun 2020 di Universitas Muhammadiyah Kalimantan Timur yang diteliti oleh (Verdikha dkk., 2021), dalam penelitiannya ini menghasilkan 1648 baris dan memiliki 3 kolom yang berisikan nilai, kelas, dan jawaban esai. Nilai dari teks jawaban memiliki range “0” sampai “12” dari jumlah kelas “A” sampai “H” dengan keseluruhan memperoleh data sebanyak 1648 baris.

```

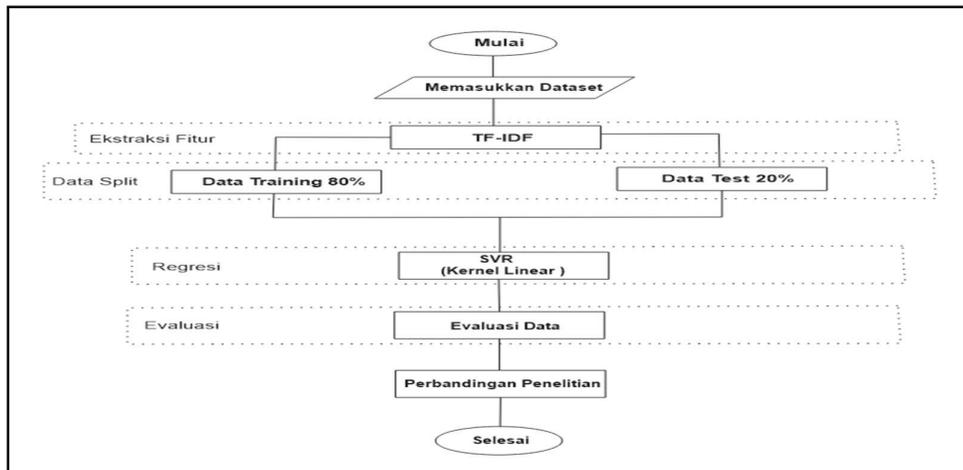
nilai kelas jawaban
8,A, 4 faktor yg sebab sebab bahasa melayu angkat j...
8,A, bahasa melayu angkat jadi bahasa satu indonesi...
8,A, empat faktor sebab bahasa melayu angkat jadi b...
8,A, alas bahasa melayu pilih jadi bahasa indonesia...
6,A, bahasa indonesia tumbuh kembang bahasa melayu ...
...
6,H, bahasa rupa salah satu unsur identitas suatu b...
4,H, cakup jumlah bahasa saling mirip tutur wilayah...
4,H, empat faktor bahasa melayu angkat jadi bahasa ...
4,H, memang banyak guna bagi besar masyarakat indon...
4,H, bahasa melayu ini akhir jadi bahasa indonesia ...
  
```

Gambar 2 Memasukkan Data

Data diatas sebelumnya sudah melalui beberapa tahapan-tahapan pra-proses. Tujuannya agar data lebih mudah diproses dengan menghilangkan beberapa kata imbuhan dan tanda simbol pada jawaban, sehingga pada saat memproses data ke dalam ekstraksi fitur, pembobotannya dapat lebih baik.

2.6 Tahapan Penelitian

Tahap penelitian ini akan mencari nilai RMSE menggunakan metode SVR dengan *kernel linear* dengan data jawaban siswa sebanyak 1648 dan membandingkan hasil tersebut dengan penelitian sebelumnya yang dilakukan oleh (Verdikha dkk., 2021), yang mana dalam penelitian Verdikha menggunakan SVR dengan *kernel RBF*. Adapun *flowchart* yang dibuat untuk mempermudah berjalannya konsep penelitian ini bisa dilihat pada gambar 3 dibawah.



Gambar 3 Tahapan Penelitian

Gambar 3 di atas menjelaskan tahapan-tahapan untuk mendapatkan nilai evaluasi RMSE agar dapat melakukan perbandingan nilai SVR *kernel liniear* dengan SVR *kernel RBF*. Berdasarkan pada gambar 2 langkah pertama dalam penelitian yaitu dimulai dengan memasukan dataset, langkah kedua dilakukan ekstraksi fitur dengan TF-IDF, kemudian langkah ketiga melakukan split data dengan 80% data training dan 20 data testing, tahap berikutnya dilakukan regresi menggunakan kernel linier, kemudian dilakukan evaluasi data, pada tahap akhir dilakukan perbandingan penelitian. Adapun bahasa pemrograman yang digunakan pada penelitian ini adalah *python* dengan versi 3.8.3 dan *tools* yang digunakan adalah *jupyter notebook* versi 6.4.5 dan modul *library* yang digunakan untuk melakukan *regresi* dan evaluasi pada penelitian Barupal & Fiehn (Barupal & Fiehn, 2019) adalah *sklearn* versi 1.1.1.

3. Hasil Penelitian

3.1 Hasil Memasukan Data

Pada gambar 4 dibawah, *dataset* mempunyai 1648 baris dan 3 kolom, dimana 1648 baris ini adalah jumlah data dan 3 kolom adalah jumlah kolom, dimana masing-masing kolomnya memiliki atribut yaitu nilai, kelas, dan jawaban.

	nilai	kelas	jawaban
0	8	A	4 faktor yg sebab sebab bahasa melayu angkat j...
1	8	A	bahasa melayu angkat jadi bahasa satu indonesi...
2	8	A	empat faktor sebab bahasa melayu angkat jadi b...
3	8	A	alas bahasa melayu plilh jadi bahasa Indonesia...
4	6	A	bahasa Indonesia tumbuh kembang bahasa melayu ...
...
1643	6	H	bahasa rupa salah satu unsur identitas suatu b...
1644	4	H	cakup jumlah bahasa saling mirip tutur wilayah...
1645	4	H	empat faktor bahasa melayu angkat jadi bahasa ...
1646	4	H	memang banyak guna bagi besar masyarakat indon...
1647	4	H	bahasa melayu ini akhir jadi bahasa indonesia ...

1648 rows x 3 columns

Gambar 4 Hasil Memasukan Data

Agar dapat mengetahui rincian Isi dari data nilai yang ada pada gambar 4 bisa kita lihat pada tabel 1 dibawah sebagai berikut.

Tabel 1 Distribusi Data

DATA	
Nilai	Jumlah
0	29
1	4
2	173
3	22
4	185
5	131
6	279
7	149
8	333
9	135
10	138
11	18
12	52
Total	1648

Tabel 1 diatas menunjukkan bahwa data dari nilai 0 adalah 29 dan nilai 1 berjumlah 4 dan begitu seterusnya hingga mendapatkan hasil keseluruhannya yaitu 1648 seperti yang terlihat pada tabel diatas. Untuk kode yang ada pada gambar 5 fungsinya adalah sebagai pemetaan antara variabel X dan y, adapun hasil outputan dari variabel X dapat dilihat pada gambar 6.

```

1 X = df1.iloc[:,2]
2 y = df1.iloc[:,0]
3
4 print(X)
5 print(y)

```

Gambar 5 Pemetaan antara Variabel X dan y

```

0      4 faktor yg sebab sebab bahasa melayu angkat j...
1      bahasa melayu angkat jadi bahasa satu indonesi...
2      empat faktor sebab bahasa melayu angkat jadi b...
3      alas bahasa melayu pilih jadi bahasa indonesia...
4      bahasa indonesia tumbuh kembang bahasa melayu ...
...
1643   bahasa rupa salah satu unsur identitas suatu b...
1644   cakup jumlah bahasa saling mirip tutur wilayah...
1645   empat faktor bahasa melayu angkat jadi bahasa ...
1646   memang banyak guna bagi besar masyarakat indon...
1647   bahasa melayu ini akhir jadi bahasa indonesia ...
Name: jawaban, Length: 1648, dtype: object

```

Gambar 6 Data dari Variabel X

Gambar 6 menjelaskan bahwa variabel X sendiri merupakan atribut dari data teks jawaban pada *dataset* sebanyak 1648 dengan *type* data object. Adapun variabel y yang jika dijalankan maka hasil outputnya dapat dilihat pada gambar 7.

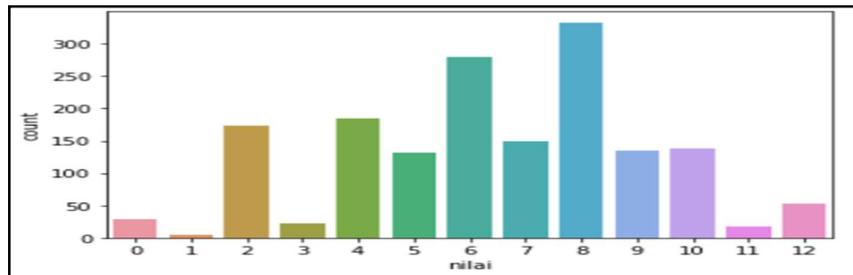
```

0      8
1      8
2      8
3      8
4      6
..
1643   6
1644   4
1645   4
1646   4
1647   4
Name: nilai, Length: 1648, dtype: int64

```

Gambar 7 Data Variabel Y

Adapun isi dari variabel y yang ada pada gambar 7 diatas adalah atribut nilai pada masing-masing teks jawaban sebanyak 1648 dengan tipe data int64. Data di atas akan dimasukkan kedalam sebuah diagram plot agar mempermudah dalam melakukan analisis nilai, jika program dijalankan maka grafiknya seperti gambar 8.



Gambar 8 Jumlah Nilai Y dalam Bentuk Diagram

Tabel 2 dibawah menunjukkan jumlah dari *Data training* 80% yang mana dari nilai ke 0 sampai dengan 12 memiliki jumlah sebanyak 1318, dan nilai *data test* 20% sebanyak 330, selanjutnya *data training* 70% yang mana dari nilai ke 0 sampai dengan 12 memiliki jumlah sebanyak 1153, dan nilai dari *data test* 30% sebanyak 495, dan yang terakhir adalah *data training* 60% dari nilai ke-0 sampai dengan 12 memiliki jumlah sebanyak 988 dan untuk *data test* 40% sebanyak 655 hasil dari perhitungan jumlah dari *data training* dan *data test* diatas dihitung menggunakan *modul len*.

Tabel 2 Hasil Data split

Data Split						
Nilai	Data training			Data test		
	80%	70%	60%	20%	30%	40%
0	25	21	17	4	8	12
1	3	2	2	1	2	2
2	145	125	17	28	4	5
3	19	18	17	3	4	113
4	146	129	110	39	56	75
5	98	83	73	33	48	58
6	227	197	166	52	82	133
7	129	113	101	20	36	48
8	257	231	201	76	102	132
9	109	90	76	26	45	59
10	104	92	78	34	46	60
11	15	12	9	3	6	9
12	41	40	29	11	12	23
Total	1316	1153	988	330	495	655

1. *Data training*

Tahapan selanjutnya adalah hasil Output *X_train* dan *y_train*, pada *data training* 80% dan *data test* 20%. Hasil output *X_train* adalah data latih pada jawaban mahasiswa yang ada pada lampiran 5 baris ke 9 dan bisa dilihat hasilnya pada gambar 12 dibawah, terlihat bahwa (0, 1855) disini menunjukkan bahwa pada index ke 0 terdapat fitur ke 1855 dengan nilai 0,047 sampai pada (1317, 254) merupakan index ke 1317 terdapat fitur 254 dengan nilai 0,019, nilai ini didapatkan dari hasil nilai ekstraksi fitur menggunakan TF-IDF yang telah di Normalisasi.

(0, 1855)	0.04773434328433336
(0, 1848)	0.16106234101737638
(0, 1832)	0.06209727701510123
(0, 1825)	0.07203613270530616
(0, 1815)	0.0800382441351015
.....
(1317, 1154)	0.3051426627544617
(1317, 748)	0.2832444951734622
(1317, 556)	0.8167499341646405
(1317, 436)	0.346943409989613
(1317, 254)	0.19801597568853432

Gambar 12 Hasil Output X_train 80:20

584	11
63	2
1346	8
1639	4
1050	8
.	.
1130	4
1294	10
860	8
1459	5
1126	2
Name: nilai, Length: 1318, dtype: int64	

Gambar 13 Hasil Output y_train 80:20

Gambar 13 diatas merupakan hasil output pada data latih *y_train*, yang mana data tersebut adalah data latih nilai siswa yang jumlah datanya sebanyak 1318, pada gambar 4.9 diatas menunjukkan bahwa pada dokumen ke-584 terdapat data fitur yang bernilai 11, sampai pada dokumen ke-1126 terdapat data fitur yang bernilai 2.

2. Data test

Tahapan selanjutnya adalah hasil Output *X_test* dan *y_test* pada *data training* 80% dan *data test* 20%. Gambar 14 dibawah menunjukkan hasil output dari *X_test*, terlihat pada (0,1735) disitu menunjukan bahwa pada index ke 0 terdapat fitur ke-1735 yang memiliki nilai sebesar 0,122 sampai pada index ke-329 yang terdapat fitur ke-19 yang memiliki nilai sebesar 0,102.

```
(0, 1735)    0.12249928128199475
(0, 1641)    0.2319556171493889
(0, 1597)    0.1202411050840897
(0, 1516)    0.0999066988576344
(0, 1244)    0.1701899107224009
.....
(329, 170)   0.10297774311911342
(329, 130)   0.07172945443204456
(329, 125)   0.10235597232902427
(329, 45)    0.10031429773725271
(329, 19)    0.10285258014362524
```

Gambar 14 Hasil Output *X_test* 80:20

```
682         6
1318        2
576         8
1636        2
331         8
.....
1313        10
1296        4
306         10
192         6
309         10
Name: nilai, Length: 330, dtype: int64
```

Gambar 15 Hasil Output *y_test* 80:20

Gambar 15 diatas merupakan output dari *data test y_test* atau *data test* nilai siswa yang menunjukkan bahwa pada dokumen ke-682 terdapat data fitur yang bernilai 6, sampai pada dokumen ke-309 terdapat data fitur yang bernilai 10.. Untuk hasil outputan *X_train*, *X_test* dan *y_train*, *y_test* dari *data split* 70:30 dan 60:40 dapat dilihat pada gambar dibawah berikut.

```
(0, 1755)    0.149611768805108725
(0, 1641)    0.17508838163823368
(0, 1516)    0.15082628679378512
(0, 1110)    0.27062468370277415
.....
(1152, 1154) 0.3051426627544617
(1152, 748)  0.2832444951734622
(1152, 856)  0.8167499341646409
(1152, 436)  0.346943409989613
(1152, 254)  0.19801597568853432
.....
285         8
1304        2
428         5
1608        4
757         10
.....
1130        4
1294        10
860         8
1459        5
1126        2
Name: nilai, Length: 1153, dtype: int64
```

Gambar 16 Hasil Output *y_train* 70:30

```

(0, 1735) 0.12249928128199475
(0, 1641) 0.2319556171493889
(0, 1597) 0.1202411050840897
(0, 1516) 0.0999066988576344
(0, 1244) 0.1701899107224009
.....
(494, 436) 0.09067355845895082
(494, 365) 0.09211651338095676
(494, 254) 0.6727684233091344
(494, 213) 0.11048499897502537
(494, 130) 0.09461682032680302

```

```

682 6
1318 2
576 8
1636 2
331 8
.....
1258 6
1283 6
18 10
423 2
1133 8
Name: nilai, Length: 495, dtype: int64

```

Gambar 17 Hasil Output X_{test} 70:30

```

(0, 1842) 0.11698645042662435
(0, 1755) 0.1331533958295138
(0, 1746) 0.1022774635895546
(0, 1687) 0.18830499970878153
(0, 1582) 0.18830499970878153
.....
(987, 1154) 0.3051426627544617
(987, 748) 0.2832444951734622
(987, 556) 0.8167499341646405
(987, 436) 0.346943409989613
(987, 254) 0.19801597568853432

```

```

1562 4
1524 8
679 7
328 8
772 9
.....
1130 4
1294 10
860 8
1459 5
1126 2
Name: nilai, Length: 988, dtype: int64

```

Gambar 18 Hasil Output y_{train} 60:40

```

(0, 1735) 0.12249928128199475
(0, 1641) 0.2319556171493889
(0, 1597) 0.1202411050840897
(0, 1516) 0.0999066988576344
(0, 1244) 0.1701899107224009
.....
(659, 369) 0.32413601038527584
(659, 254) 0.26687150559320877
(659, 175) 0.11758737797869831
(659, 72) 0.16151314638990236
(659, 43) 0.160747510153808

```

```

682 6
1318 2
576 8
1636 2
331 8
.....
333 8
1232 8
1447 2
1307 2
889 5
Name: nilai, Length: 660, dtype: int64

```

Gambar 19 Hasil Output X_{test} 60:40

3.4 Regresi SVR Kernel (“Linear & Rbf”)

Tahapan selanjutnya merupakan hasil dari melakukan prediksi data nilai *test* dan *training*, yang mana hasil dari melakukan redidasi dan tahap-tahapannya sama seperti penelitian terdahulu yang dilakukan oleh (Verdikha dkk., 2021), Bisa dilihat pada gambar 20, gambar tersebut merupakan hasil dari data yang diprediksi sebanyak 330 sesuai dengan *data test* 20%, adapun hasil dari *regresi data test* dapat dilihat pada gambar 20 dibawah.

Jumlah data yang di Prediksi dari Data Test 20% = 330					
5.26624322	2.97829206	6.49180642	4.00273028	6.41946493	8.61879939
6.91847394	6.2933744	4.78593464	4.18500647	4.86246342	5.48549697
6.53602807	7.87482822	5.97176068	5.69812596	7.19261522	7.09968175
7.65750987	8.7297767	4.45198588	6.67616626	4.20283617	5.27362663
6.52744612	4.89963091	7.19261522	5.39806002	4.19753779	6.14628617
.....
6.29037866	5.30528039	6.04283755	4.31828422	7.22988262	5.9578466
6.38068731	7.09968175	6.07757746	4.56075541	6.40167168	7.98132153
7.65750987	7.06683135	9.12013437	4.92437362	5.73859511	4.27526056
6.09999088	8.9253575	7.36908408	9.12013437	5.61368905	7.09968175
6.09609688	9.53519688	6.50016325	8.8999774	6.69449304	6.88697054

Gambar 20 Nilai Prediksi Data test 20%

Pada doc ke = 682 data yang di prediksi SVR kernel linear = 5.2662432162604995 dan nilai aktualnya adalah 6
Pada doc ke = 1318 data yang di prediksi SVR kernel linear = 2.9782920571406817 dan nilai aktualnya adalah 2
Pada doc ke = 576 data yang di prediksi SVR kernel linear = 6.491806416585913 dan nilai aktualnya adalah 8
Pada doc ke = 1636 data yang di prediksi SVR kernel linear = 4.002730276526617 dan nilai aktualnya adalah 2
Pada doc ke = 331 data yang di prediksi SVR kernel linear = 6.419464931548707 dan nilai aktualnya adalah 8
Pada doc ke = 903 data yang di prediksi SVR kernel linear = 8.618799385417764 dan nilai aktualnya adalah 5
Pada doc ke = 173 data yang di prediksi SVR kernel linear = 6.91847393547547 dan nilai aktualnya adalah 9
Pada doc ke = 829 data yang di prediksi SVR kernel linear = 6.293374396588537 dan nilai aktualnya adalah 4
Pada doc ke = 493 data yang di prediksi SVR kernel linear = 4.785934639274888 dan nilai aktualnya adalah 6
Pada doc ke = 374 data yang di prediksi SVR kernel linear = 4.185006472044927 dan nilai aktualnya adalah 1

Gambar 21 Nilai Prediksi dan Nilai Aktual SVR Kernel Linear

Sebagai contoh rincian dari data pada gambar 21 diatas adalah hasil prediksi menggunakan parameter *kernel linear* untuk baris pertama dan kolom pertama hasil prediksi dari nilai teks jawaban baris ke 682 memiliki nilai prediksi SVR *kernel linear* sebesar 5,266 dengan nilai aktualnya adalah 6, untuk jawaban baris ke 1318 memiliki nilai prediksi SVR *kernel linear* 2,978, sedangkan nilai aktualnya adalah 2. Untuk hasil dari *regresi data test* 30% dan 40% menggunakan metode SVR *kernel linear*, hasilnya dapat dilihat pada gambar 22 - 25.

Jumlah data yang di Prediksi dari Data Test 30% = 495					
5.05798369	3.13091829	6.51744573	3.71325951	6.6006347	8.58411306
6.99181731	6.29840942	4.71721487	4.31571742	4.89093408	5.75567829
6.71755268	7.65729744	6.23605076	6.0915353	7.22410597	7.04122968
7.72464959	8.63298672	4.63641075	6.62680244	4.99714501	5.10473764
6.61814913	4.89981533	7.22410597	5.30104315	4.14112182	5.94224727
.....
7.07626442	7.32462682	4.45742952	5.84201817	8.20040923	6.50006342
6.80427638	6.37198764	5.96936613	7.33809096	5.56089482	5.34862761
5.66265985	2.78751391	5.25265052	3.5991515	7.22410597	7.22410597
8.95104359	5.95768349	7.45387686	7.93814588	7.69796309	4.27451997
5.53253175	5.28923541	9.11623129	3.91429181	6.53152889	9.02403413
7.70577779	3.87144767	7.22410597			

Gambar 22 Hasil Regresi Data Test 30% dengan Kernel Linear

Pada doc ke = 682 data yang di prediksi SVR kernel linear = 5.05798368707009 dan nilai aktualnya adalah 6
Pada doc ke = 1318 data yang di prediksi SVR kernel linear = 3.130918286174151 dan nilai aktualnya adalah 2
Pada doc ke = 576 data yang di prediksi SVR kernel linear = 6.517445729115684 dan nilai aktualnya adalah 8
Pada doc ke = 1636 data yang di prediksi SVR kernel linear = 3.7132595056055444 dan nilai aktualnya adalah 2
Pada doc ke = 331 data yang di prediksi SVR kernel linear = 6.6006346950222845 dan nilai aktualnya adalah 8
Pada doc ke = 903 data yang di prediksi SVR kernel linear = 8.58411306204574 dan nilai aktualnya adalah 5
Pada doc ke = 173 data yang di prediksi SVR kernel linear = 6.991817311229225 dan nilai aktualnya adalah 9

Gambar 23 Nilai Prediksi dan Nilai Aktual SVR Kernerl Linier

Jumlah data yang di Prediksi dari Data Test 40% = 655					
[5.27671023	3.07359747	6.51575847	3.56249203	6.73382333	8.34787816
7.24486871	6.06369265	4.94542617	4.11882141	4.49167191	5.94944315
6.65919772	8.00625601	6.65351819	6.54527631	7.42082079	7.0395148
7.79325506	8.42311617	4.48423819	7.00608009	5.10217144	5.18450478
6.6494152	4.89992562	7.42082079	5.5005527	4.34184364	6.12217149
.....
4.32178318	6.32381669	4.54178966	8.09998136	5.58836374	9.12887369
8.40112217	4.8471184	5.3949885	7.10048672	5.08943806	5.55363626
8.9653531	7.63337272	7.86759103	6.94289347	8.11659767	8.42581531
8.10132534	6.5360859	5.8649587	6.74214486	8.04512463	6.61189652
8.75831601	4.26368123	4.95732477	4.32178318	3.32654117	7.3750335]

Gambar 24 Hasil Regresi Data Test 40% dengan Kernel Linier

Pada doc ke = 682 data prediksi SVR kernel Linear = 5.276710232851965 dan nilai aktualnya adalah 6
Pada doc ke = 1318 data prediksi SVR kernel Linear = 3.0735974716624344 dan nilai aktualnya adalah 2
Pada doc ke = 576 data prediksi SVR kernel Linear = 6.515758465837244 dan nilai aktualnya adalah 8
Pada doc ke = 1636 data prediksi SVR kernel Linear = 3.5624920300265375 dan nilai aktualnya adalah 2
Pada doc ke = 331 data prediksi SVR kernel Linear = 6.73382333221188 dan nilai aktualnya adalah 8
Pada doc ke = 903 data prediksi SVR kernel Linear = 8.347878163716603 dan nilai aktualnya adalah 5
Pada doc ke = 173 data prediksi SVR kernel Linear = 7.2448687147940305 dan nilai aktualnya adalah 9

Gambar 25 Nilai Prediksi dan Nilai Aktual SVR Kernerl Linier

Selanjutnya peneliti menampilkan hasil dari penerapan *regresi* menggunakan algoritma SVR dengan *kernel* RBF menggunakan *data test* 20%, yang mana hasilnya dapat dilihat pada gambar 26.

Jumlah data yang di Prediksi dari Data Test 20% = 330					
[5.59928761	4.38245974	6.47877673	4.75735229	5.93164435	8.86540534
6.7700525	6.19942101	5.16223632	4.5232494	5.17992952	5.34362661
6.77294936	7.55285583	6.18012135	5.82299298	7.09967045	7.56080905
7.68757503	9.06178944	4.45874976	6.73553717	4.53336562	5.0323958
7.28880881	4.90049893	7.09967045	5.29532414	4.26250253	6.08890607
.....
6.39541489	5.29729622	6.13353021	4.35819098	6.79801557	6.13765607
6.5803586	7.56080905	5.92603181	5.26768358	6.02028495	7.90434123
7.68757503	6.79648303	9.10039317	4.77954419	6.05012251	3.91536174
6.27758294	9.41738781	8.09384299	9.10039317	5.7256785	7.56080905
6.24987579	9.10002774	6.58103983	8.90001715	6.47929825	7.25613553]

Gambar 26 Nilai Prediksi Data Test 20%

Pada doc ke = 682 data yang di prediksi SVR kernel rbf = 5.59928761248673 dan nilai aktualnya adalah 6
Pada doc ke = 1318 data yang di prediksi SVR kernel rbf = 4.382459735298608 dan nilai aktualnya adalah 2
Pada doc ke = 576 data yang di prediksi SVR kernel rbf = 6.478776725063636 dan nilai aktualnya adalah 8
Pada doc ke = 1636 data yang di prediksi SVR kernel rbf = 4.7573522930853205 dan nilai aktualnya adalah 2
Pada doc ke = 331 data yang di prediksi SVR kernel rbf = 5.931644350618654 dan nilai aktualnya adalah 8
Pada doc ke = 903 data yang di prediksi SVR kernel rbf = 8.865405337798975 dan nilai aktualnya adalah 5
Pada doc ke = 173 data yang di prediksi SVR kernel rbf = 6.7700524973959935 dan nilai aktualnya adalah 9
Pada doc ke = 829 data yang di prediksi SVR kernel rbf = 6.19942100858604 dan nilai aktualnya adalah 4
Pada doc ke = 493 data yang di prediksi SVR kernel rbf = 5.162236320473932 dan nilai aktualnya adalah 6
Pada doc ke = 374 data yang di prediksi SVR kernel rbf = 4.523249395324046 dan nilai aktualnya adalah 1

Gambar 27 Nilai Prediksi dan Nilai Aktual SVR Kernel RBF

Pada gambar 27 hasil prediksi nilai menggunakan *kernel* rbf. untuk baris pertama dan kolom pertama hasil prediksi dari nilai teks jawaban baris ke-682 memiliki nilai prediksi dengan SVR *kernel* rbf sebesar 5,599 dan nilai aktual nya 6, dan kedua, hasil dari nilai teks jawaban baris ke-1318 memiliki nilai aktual sebesar 2. Untuk hasil dari *regresi data test* 30% dan 40% menggunakan metode SVR *kernel* rbf, hasilnya dapat dilihat pada gambar 28 – 31.

Jumlah data yang di Prediksi dari Data Test 30% = 495					
5.43534664	4.54598445	6.46457742	4.73415279	6.10333312	8.75713365
6.96284459	6.2457446	5.30111416	4.66748047	5.20174211	5.53201073
6.8779721	7.45258608	6.32783098	5.98436024	7.09964594	7.68852986
7.82879342	8.89981764	4.58413512	6.71580718	5.35718055	4.86490359
6.97834728	4.89992504	7.09964594	5.30635056	4.2415362	6.03310144
.....
6.6898811	6.14306487	6.09362502	7.51274951	5.68731032	5.59833543
5.92202245	4.04746044	5.75961627	4.35554545	7.09964594	7.09964594
8.63179117	5.82233233	6.9011371	7.89421366	7.69316345	4.40301542
6.03825738	5.98242715	9.06454356	3.87444695	5.770393	8.46100096
7.43955685	5.09506786	7.09964594			

Gambar 28 Nilai Data Test 30% dengan SVR Kernel RBF

Pada doc ke = 682 data yang di prediksi SVR kernel linear = 5.4353466414105345 dan nilai aktualnya adalah 6
Pada doc ke = 1318 data yang di prediksi SVR kernel linear = 4.545984450824466 dan nilai aktualnya adalah 2
Pada doc ke = 576 data yang di prediksi SVR kernel linear = 6.464577421425391 dan nilai aktualnya adalah 8
Pada doc ke = 1636 data yang di prediksi SVR kernel linear = 4.734152786050478 dan nilai aktualnya adalah 2
Pada doc ke = 331 data yang di prediksi SVR kernel linear = 6.103333122032971 dan nilai aktualnya adalah 8
Pada doc ke = 903 data yang di prediksi SVR kernel linear = 8.75713364676333 dan nilai aktualnya adalah 5
Pada doc ke = 173 data yang di prediksi SVR kernel linear = 6.962844587671119 dan nilai aktualnya adalah 9

Gambar 29 Nilai Prediksi dan Nilai Aktual SVR Kernel RBF

Jumlah data yang di Prediksi dari Data Test 40% = 655					
5.47746916	4.42001301	6.42956631	4.60847224	6.22006196	8.67025127
7.23253862	5.7200164	5.17593737	4.59431307	5.0504945	5.54457034
6.81352404	7.62932501	6.63832261	6.06690734	7.41586245	7.70158877
7.86664975	8.86614902	4.501705	6.93210589	5.28206701	4.92504427
7.19523777	4.89977451	7.41586245	5.37443353	4.56010025	5.98388071
.....
3.63684705	5.99483716	4.99637413	8.09993407	5.6325012	9.01159996
7.70001712	5.04826034	5.18110996	7.10017172	5.51599159	5.727216
6.64138413	7.23737763	7.72258826	6.65297142	7.42740204	8.84631746
6.84000136	6.05967645	5.65540578	6.51369105	8.04453326	7.05361038
8.42222531	4.36841922	5.35410971	3.63684705	3.63252016	6.61690713

Gambar 30 Nilai Data Test 40% dengan SVR Kernel RBF

Pada doc ke = 682 data prediksi SVR kernel Linear = 5.477469161861146 dan nilai aktualnya adalah 6
Pada doc ke = 1318 data prediksi SVR kernel Linear = 4.42001301432293 dan nilai aktualnya adalah 2
Pada doc ke = 576 data prediksi SVR kernel Linear = 6.42956630651474 dan nilai aktualnya adalah 8
Pada doc ke = 1636 data prediksi SVR kernel Linear = 4.608472243821204 dan nilai aktualnya adalah 2
Pada doc ke = 331 data prediksi SVR kernel Linear = 6.220061961758571 dan nilai aktualnya adalah 8
Pada doc ke = 903 data prediksi SVR kernel Linear = 8.670251273181929 dan nilai aktualnya adalah 5
Pada doc ke = 173 data prediksi SVR kernel Linear = 7.232538616091249 dan nilai aktualnya adalah 9

Gambar 31 Nilai Prediksi dan Nilai Aktual SVR Kernel RBF

3.5 Evaluasi RMSE

Tahapan evaluasi berguna untuk menghitung akar kuadrat dari nilai prediksi dan nilai aktual yang sudah di *regresi* sebelumnya menggunakan metode SVR, hasil evaluasi RMSE menggunakan metode SVR dengan *kernel linier* dan *kernel rbf* dapat dilihat pada tabel 3.

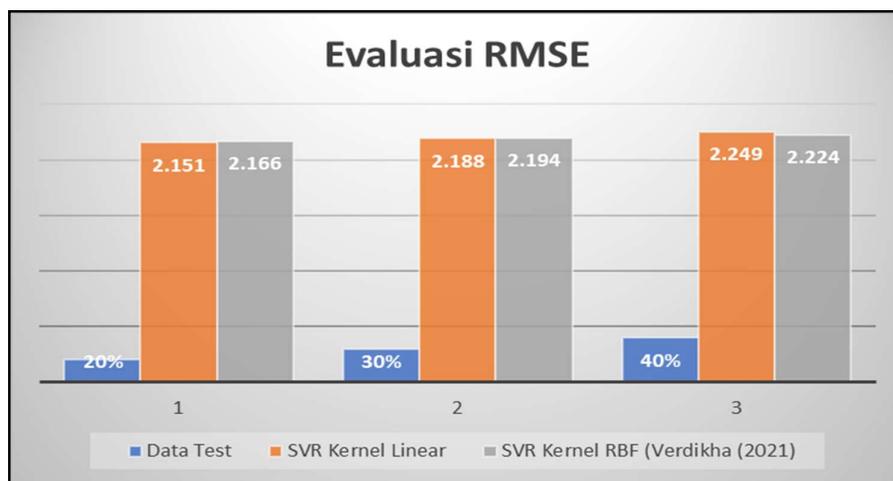
Tabel 3 Hasil Perbandingan SVR

Data test	20%	30%	40%
SVR Kernel linier	2.151	2.188	2.249
SVR Kernel RBF (Verdikha (2021))	2.166	2.194	2.224
Rentang Perbedaan	0.015	0.006	0.025

Tabel 3 diatas menjelaskan hasil dari evaluasi RMSE menggunakan metode SVR dengan *kernel liniear* dan *kernel rbf*, yang mana pada *data test* 20% dari SVR dengan *kernel liniear* mendapatkan hasil evaluasi RMSE sebesar 2,151 dan SVR dengan *kernel rbf* sebesar 2,166 yang memiliki rentang perbedaan sebesar 0,015, untuk *data test* 30% dari SVR dengan *kernel liniear* mendapatkan hasil evaluasi RMSE sebesar 2,188 dan SVR dengan *kernel rbf* sebesar 2,194 yang memiliki rentang perbedaan sebesar 0,006, dan yang terakhir adalah *data test* 40% dari SVR dengan *kernel liniear* mendapatkan hasil evaluasi RMSE sebesar 2,249 dan SVR dengan *kernel rbf* sebesar 2,224 yang memiliki rentang perbedaan sebesar 0,025.

3.6 Hasil Perbandingan

Perbandingan dari hasil evaluasi antara *kernel liniear* dengan penelitian sebelumnya yang dilakukan oleh (Verdikha 2021) yaitu *kernel RBF* dapat dilihat pada diagram 4.18 dibawah.



Gambar 32 Diagram Perbandingan Evaluasi RMSE

Gambar 32 di atas menjelaskan bahwa, penelitian menggunakan *data test* 20% dan 30% dengan SVR *kernel liniear* mendapatkan hasil evaluasi RMSE lebih kecil dari pada SVR dengan *kernel rbf*, yang mana hasil dari evaluasi menggunakan SVR *kernel liniear* sebesar 2,151 untuk *data test* 20% dan untuk *data test* 30% sebesar 2,188 dan hasil evaluasi menggunakan SVR *kernel rbf* sebesar 2,166 untuk *data test* 20% dan untuk *data test* 30% sebesar 2,194, grafik diatas menunjukkan evaluasi *data test* 20% dan 30% yang terbaik ada pada *kernel liniear*, yang mana bisa dilihat pada hasil prediksi dan nilai aktual antara dua penelitian yang bisa dilihat pada gambar 4.13 dan gambar 4.15 yang mana hasil dari nilai *test* jawaban baris ke-682 sudah menunjukkan hasil yang jauh berbeda, yaitu nilai prediksi dari SVR *kernel liniear* mendapatkan hasil sebesar 5,266 dan *kernel rbf* sebesar 5,599, yang memiliki selisih sebesar 0,333.

Perbandingan hasil nilai evaluasi RMSE pada penelitian ini dengan penelitian sebelumnya memiliki selisih 0,015, hasil evaluasi RMSE pada penelitian sebelumnya yang dilakukan oleh Verdikha et al (2021), yang menggunakan metode SVR dengan *kernel RBF* mendapatkan hasil nilai evaluasi RMSE sebesar 2,166, dan pada penelitian ini yang menggunakan metode SVR dengan *kernel liniear* mendapatkan hasil nilai evaluasi RMSE sebesar 2,151. Sehingga membuktikan bahwa penggunaan *kernel liniear* dalam penggunaan metode *regresi SVR* lebih baik dari *kernel RBF*.

4. Kesimpulan

Pada penelitian ini hasil evaluasi menggunakan svr dengan *kernel liniear* mendapatkan nilai RMSE sebesar 2,151. Diketahui bahwa pembuatan model

pembelajaran *Auto Essay Scoring* dengan ekstraksi fitur *TF-IDF* dan melakukan *data split* dengan rasio 80:20 menggunakan algoritma *Support Vector Regression* dengan *kernel linier* telah berhasil dan menghasilkan evaluasi *RMSE* sebesar 2,151 dengan selisih evaluasi *RMSE* dengan penelitian sebelumnya yang dilakukan oleh (Verdikha et al., 2021) dengan menggunakan *kernel RBF* sebesar 0,015, Sehingga membuktikan bahwa penggunaan *kernel linier* dalam penggunaan metode *regresi SVR* lebih baik dari *kernel RBF*.

Daftar Referensi

- [1] Asyiva, A. (2019). Prediksi Laju Inflasi Indonesia Menggunakan Metode Support Vector Regression Dengan Kernel Radial Basis Function. *Repository.Uinjkt.Ac.Id*.
- [2] Barupal, D. K., & Fiehn, O. (2019). Generating the blood exposome database using a comprehensive text mining and database fusion approach. *Environmental Health Perspectives*, 127(9), 2825–2830. <https://doi.org/10.1289/EHP4713>
- [3] Cholis, M. N., Yudaningtyas, E., & Aswin, M. (2019). Pengaruh Penggunaan Synonym Recognition dan Spelling Correction pada Hasil Aplikasi Penilaian Esai dengan Metode Longest Common Subsequence dan Cosine Similarity. *InfoTekJar (Jurnal Nasional Informatika dan Teknologi Jaringan)*, 3(2), 138–142. <https://doi.org/10.30743/infotekjar.v3i2.1061>
- [4] Dalimunthe, K., & Hayadi, B. H. (2022). Information Text Retrieval Untuk Pencarian Data Penilaian Mengacu Pada Saran Dari Pengunjung Menggunakan Vector Space Model. *Journal Computer Science and Information Technology(JCoiNT)*, 3(2), 73–79. <https://jurnal.ulb.ac.id/index.php/JCoiNT/article/view/3110>
- [5] Julianto, Y., Setiabudi, D. H., Rostianingsih, S., Informatika, P. S., Petra, U. K., & Surabaya, J. S. (2022). *Analisis Sentimen Ulasan Restoran Menggunakan Metode Support Vector Machine*.
- [6] Lestari, M. P., Witarsyah, D. J., Hamami, F., Telkom, U., Regression, S. V., Absolute, M., & Error, P. (2021). PERAMALAN PERTAMBAHAN PASIEN COVID-19 MENGGUNAKAN SUPPORT VECTOR REGRESSION FORECASTING GROWTH OF COVID-19 PATIENTS USING SUPPORT. 8(5), 9497–9507.
- [7] Maulana, N. D., Setiawan, B. D., & Dewi, C. (2019). Implementasi Metode Support Vector Regression (SVR) Dalam Peramalan Penjualan Roti (Studi Kasus : Harum Bakery). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(3), 2986–2995.
- [8] Maulidya, C. M. (2019). Program studi informatika fakultas teknik universitas 17 agustus 1945 surabaya 2019. *Repository.Untag-Sby.Ac.Id*, 1411506486. <http://repository.untag-sby.ac.id/12136/1/1461700227-Rudolf-UTS.pdf>
- [9] Pangarkar, D. J., Sharma, R., Sharma, A., & Sharma, M. (2020). Assessment of the Different Machine Learning Models for Prediction of Cluster Bean (*Cyamopsis tetragonoloba* L. Taub.) Yield. *Advances in Research, August*, 98–105. <https://doi.org/10.9734/air/2020/v21i930238>
- [10] Ramesh, D., & Sanampudi, S. K. (2021). An automated essay scoring systems: a systematic literature review. In *Artificial Intelligence Review* (Nomor 0123456789). Springer Netherlands. <https://doi.org/10.1007/s10462-021-10068-2>
- [11] Thamrin, H., Verdikha, N. A., & Triyono, A. (2021). Text Classification and Similarity Algorithms in Essay Grading. *4th International Seminar on Research of Information Technology and Inteligent System (ISRITI)*.
- [12] Verdikha, N. A., Thamrin, H., Triyono, A., Fikri, M., & Suryawan, S. H. (2021). *Regression and Oversampling Method for Indonesian Language Automated Essay Scoring*. 2.

Naspub: Komparasi Evaluasi Parameter Kernel Linier Support Vector Regression (Svr) untuk Automated Essay Scoring (Aes) Bahasa Indonesia

by Riski Nopiansyah

Submission date: 06-Jun-2023 08:31AM (UTC+0800)

Submission ID: 2109873543

File name: UjiTurnitin_Naskah_Jurnal_RiskiNopiansyah_1811102441083.docx (1.44M)

Word count: 3798

Character count: 21792

Naspub: Komparasi Evaluasi Parameter Kernel Linier Support Vector Regression (Svr) untuk Automated Essay Scoring (Aes) Bahasa Indonesia

ORIGINALITY REPORT

22%	22%	7%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.umkt.ac.id Internet Source	6%
2	dspace.uui.ac.id Internet Source	2%
3	publication.petra.ac.id Internet Source	2%
4	www.scribd.com Internet Source	1%
5	repository.uinjkt.ac.id Internet Source	1%
6	repository.ub.ac.id Internet Source	1%
7	bestjournal.untad.ac.id Internet Source	1%