

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

SKRIPSI

Diajukan untuk memenuhi sebagian persyaratan mencapai gelar
Sarjana Komputer

DISUSUN OLEH :

RISKI NOPIANSYAH

1811102441083



**PROGRAM STUDI S1 TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR
SAMARINDA
2022**

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

SKRIPSI

Diajukan untuk memenuhi sebagian persyaratan mencapai gelar
Sarjana Komputer

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

DISUSUN OLEH:

RISKI NOPIANSYAH

1811102441083

Telah melaksanakan ujian skripsi dan dinyatakan lulus,

Pada tanggal 10 Desember 2022

Dosen Pembimbing



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

Penguji



Asslia Johar Latipah, M.Cs
NIDN : 1124098902

Dekan



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

Ketua Progam Studi



Asslia Johar Latipah, M.Cs
NIDN : 1124098902

PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Riski Nopiansyah

NIM : 1811102441083

Konsentrasi : Sistem Cerdas

Menyatakan bahwa Skripsi dengan judul berikut:

Komparasi Evaluasi Parameter Kernel Linear Support Vector Regression (Svr) untuk Automated Essay Scoring (Aes) Bahasa Indonesia.

Dosen Pembimbing : Naufal Azmi Verdikha, S.Kom., M.Eng

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar Sarjana Komputer, baik di Universitas Muhammadiyah Kalimantan Timur(UMKT) maupun di Perguruan Tinggi lainnya
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas Muhammadiyah Kalimantan Timur (UMKT)
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi Muhammadiyah.

Samarinda, 10 Desember 2022

Yang Menyatakan,



Riski Nopiansyah

1811102441083

PRAKATA

Dengan mengucapkan puji syukur kepada Allah SWT yang telah melimpahkan rahmat dan hidayahnya sehingga penulis dapat menyelesaikan laporan skripsi yang berjudul "*Komparasi evaluasi Parameter Kernel linier Support Vector Regression (SVR) untuk Automated essay Scoring (AES) Bahasa Indonesia*", Laporan akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana (S1) dan sebagai sarana untuk mengimplementasikan ilmu yang telah dipelajari selama masa studi di Jurusan Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Kalimantan Timur.

Penulis menyadari bahwa penyelesaian Tugas Akhir ini tidak lepas dari bimbingan dan juga bantuan dari segala pihak. Mulai dari penyusunan laporan, perancangan sistem dan pengimplementasian sistem yang mana setiap penulis memiliki kendala atau kesulitan, banyak pihak yang membantu penulis untuk menyelesaikan kendala tersebut. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan rasa terima kasih ucapan syukur kepada:

1. *Allah Subhanahu Wa Ta'ala* dan *Nabi Muhammad Shallallahu 'Alaihi Wasallam* yang telah memberi rahmat, berkah, bantuan, kekuatan, dan kemudahan sehingga Kerja Praktek ini dapat berjalan dan terselesaikan dengan baik.
2. Bapak Prof. Dr. Bambang Setiaji selaku Rektor Universitas Muhammadiyah Kalimantan Timur.
3. Bapak Prof. Ir. Sarjito, MT., Ph.D selaku Dekan Fakultas Sains Dan Teknologi Universitas Muhammadiyah Kalimantan Timur.
4. Ibu Asslia Johar Latipah, S.Kom., M.Cs selaku Ketua Program Studi Teknik Informatika Fakultas Sains Dan Teknologi Universitas Muhammadiyah Kalimantan Timur.
5. Bapak Naufal Azmi Verdhika, S.Kom., M.Eng selaku dosen pembimbing yang telah memberikan bimbingan dan masukan dalam menyusun Laporan Skripsi ini.

6. Bapak Rudiman, S.Kom., M.Sc selaku dosen wali yang telah membantu penulis dalam mengikuti dan menyelesaikan studi di Fakultas Sains dan Teknologi di Universitas Muhammadiyah Kalimantan Timur.
7. Orang tua dan keluarga yang penulis sayangi, terima kasih atas Do'a dan dukungannya baik secara moral maupun materi.
8. Teman seperjuangan Program Studi Teknik Informatika Angkatan 2018, terima kasih atas dukungan dan waktunya.

Penulis menyadari laporan ini masih perlu banyak penyempurnaan karena kesalahan dan kekurangan. Penulis terbuka terhadap kritik dan saran pembaca agar laporan ini dapat lebih baik. Apabila terdapat banyak kesalahan pada laporan ini, baik terkait penulisan maupun konten, penulis memohon maaf.

Demikian yang dapat penulis sampaikan. Akhir kata, semoga laporan skripsi ini dapat bermanfaat.

Samarinda, 22 November 2022

Yang membuat pernyataan,



Riski Nopiansyah
NIM. 1811102441083

ABSTRAK

Automated Essay Scoring (AES) merupakan sistem penilaian 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 : *Automated Essay Scoring, Natural Language Processing, Bahasa Indonesia, Support Vector Regression, Parameter kernel linier.*

ABSTRACT

Automated Essay Scoring (AES) is a system for evaluating essay exam questions in Indonesian and foreign languages 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 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; Natural Language Processing; Indonesia Language, Support Vector Regression, Linear Kernel Parameter.*

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN JUDUL	ii
HALAMAN PENGESAHAN	iii
PERNYATAAN KEASLIAN SKRIPSI	iv
PRAKATA.....	v
ABSTRAK.....	vii
<i>ABSTRACT</i>	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN	xiii
BAB 1 PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Tujuan Penelitian.....	3
BAB 2 TINJAUAN PUSTAKA.....	4
2.1. Penelitian Terkait.....	4
2.2. Teori Dasar Penelitian	5
2.2.1. <i>Natural Language Processing (NLP)</i>	5
2.2.2. <i>Automated Essay Scoring (AES)</i>	6
2.2.3. <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	6
2.2.4. <i>Support Vector Regression (SVR)</i>	7
2.2.5. <i>Kernel</i>	8
2.2.6. <i>Root Mean Square Error (RMSE)</i>	9
BAB 3 METODE PENELITIAN	12
3.1 Data	12
3.2 Tahapan Penelitian	12
3.2.1 Memasukan Data	14

2.3.2	Ekstraksi Fitur	14
3.2.3	<i>Data split</i>	15
3.2.4	<i>Regresi</i> dan Evaluasi	16
BAB 4 HASIL DAN PEMBAHASAN		19
4.1	Hasil Memasukan Data	19
4.2	Ekstraksi Data	22
4.3	<i>Data split</i>	23
4.4	<i>Regresi SVR Kernel ("Linear & Rbf")</i>	26
4.5	Evaluasi RMSE	28
4.6	Hasil Perbandingan	28
BAB 5 PENUTUP		30
5.1	Kesimpulan	30
5.2	Saran	30
LAMPIRAN		34

DAFTAR TABEL

Tabel 2.1 Referensi Penelitian	10
Tabel 3.1 Pembagian <i>Data training</i> dan <i>Data test</i>	15
Tabel 3.2 Parameter Pada Metode SVR.....	16
Tabel 4.1 Distribusi Data	19
Tabel 4.2 Hasil <i>Data split</i>	23
Tabel 4.3 Hasil Perbandingan SVR	28

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi SVR.....	7
Gambar 3.1 Memasukan Data	12
Gambar 3.2 Tahapan Penelitian.....	13
Gambar 4.1 Distribusi Data	19
Gambar 4.2 Data Dari Variabel X	20
Gambar 4.3 Data Dari Variabel Y	21
Gambar 4.4 Jumlah Nilay Y Dalam Bentuk Diagram	21
Gambar 4.5 Ekstraksi Data	22
Gambar 4.6 Nilai Hasil Normalisasi TF-IDF.....	22
Gambar 4.7 Hasil <i>Data split</i>	23
Gambar 4.8 Hasil Output <i>X_train</i> 80:20.....	24
Gambar 4.9 Hasil Output <i>Y_train</i> 80:20.....	24
Gambar 4.10 Hasil Output <i>X_Test</i> 80:20.....	25
Gambar 4.11 Hasil Output <i>Y_Test</i> 80:20.....	25
Gambar 4.12 Nilai Prediksi <i>Data test</i> 20%	26
Gambar 4.13 Nilai Prediksi Dan Nilai Aktual SVR <i>Kernel liniear</i>	26
Gambar 4.14 Nilai Prediksi <i>Data test</i> 20%	27
Gambar 4.15 Nilai Prediksi dan Nilai Aktual SVR <i>kernel</i> RBF	27
Gambar 4.16 Diagram Perbandingan Evaluasi RMSE	29

DAFTAR LAMPIRAN

- Lampiran 1. Riwayat Hidup
- Lampiran 2. Memasukan *Dataset* (data_penelitian.csv)
- Lampiran 3. Memetakan Atribut Data ke Variabel
- Lampiran 4. Menampilkan Diagram Plot Atribut Nilai
- Lampiran 5. Ekstraksi Fitur TF-IDF
- Lampiran 6. *Data split 80/20*
- Lampiran 7. *Data split 70/30*
- Lampiran 8. *Data split 60/40*
- Lampiran 9. *Regresi SVR Kernel liniear dengan data test 20%*
- Lampiran 10. *Regresi SVR Kernel liniear dengan data test 30%*
- Lampiran 11. *Regresi SVR Kernel liniear dengan data test 40%*
- Lampiran 12. Hasil *Regresi Data test 30%* dengan *Kernel liniear*
- Lampiran 13. Hasil *Regresi Data test 40%* dengan *Kernel liniear*
- Lampiran 14. Hasil *Regresi Data split 30%*
- Lampiran 15. Hasil *Regresi Data split 40%*
- Lampiran 16. Hasil Output *Data training 70:30*
- Lampiran 17. Hasil Output *Data test 70:30*
- Lampiran 18. Hasil Output *Data test 60:40*
- Lampiran 19. Hasil Output *Data test 70:30*
- Lampiran 20. Surat Ijin Penelitian
- Lampiran 21. Surat Keterangan Tidak Uji Validitas
- Lampiran 22. Lembar Bimbingan Skripsi
- Lampiran 23. Hasil Uji Plagiasi