

Implementasi Metode Support Vector Machine (SVM) dan t-Distributed Stochastic Neighbor Embedding (t-SNE) untuk Klasifikasi Depresi

Satria Abimayu¹⁾, Nurdin Bahtiar^{*2)}, Eko Adi Sarwoko³⁾

^{1,2,3} Departemen Informatika, Fakultas Sains dan Matematika, Universitas Diponegoro, Semarang
¹satriaabimayu01@gmail.com, ^{*2}nurdinbahtiar@lecturer.undip.ac.id,
³ekoadisarwoko@lecturer.undip.ac.id

Abstrak

Depresi merupakan salah satu gangguan kesehatan mental. Sekitar 300 juta jiwa atau 3,76% populasi di dunia dari segala usia dan komunitas menderita depresi. WHO memprediksi bahwa depresi akan menjadi penyebab kematian paling berdampak dalam 15 tahun ke depan. Penelitian terdahulu yang melakukan klasifikasi terhadap depresi untuk instrumen Depression Anxiety Stress Scales (DASS-42) masih sangat sedikit. Penelitian ini mengidentifikasi seseorang memiliki kemungkinan depresi, melalui proses pelatihan model klasifikasi menggunakan metode Support Vector Machine dan t-Distributed Stochastic Neighbor Embedding pada set data DASS-42. Set data DASS-42 terdiri dari 39.776 data dan dapat digunakan untuk mengklasifikasi 3 fenomena yang berbeda yaitu, depresi, stress dan kecemasan. Model Support Vector Machine dilatih menggunakan data DASS-42 yang telah dibersihkan, normalisasi dan balancing serta menggunakan atribut yang telah direduksi melalui proses reduksi dimensi t-Distributed Stochastic Neighbor Embedding. Data latih dan data uji dibagi dengan rasio 80:20. Berdasarkan hasil pengujian, implementasi metode Support Vector Machine (SVM) dan t-Distributed Stochastic Neighbor Embedding (t-SNE) untuk klasifikasi depresi pada data DASS-42 menunjukkan performa yang lebih baik dengan akurasi terbaik sebesar 100% pada data sebelum balancing dan 91,71% pada data setelah balancing.

Kata kunci : Support Vector Machine, t-Distributed Stochastic Neighbor Embedding, Klasifikasi, Depresi

Abstract

Depression is a mental health disorder. About 300 million people or 3.76% of the world's population of all ages and communities suffer from depression. WHO predicts that depression will be the most impactful cause of death in the next 15 years. Very few previous studies have classified depression for the Depression Anxiety Stress Scales (DASS-42) instrument. This study identifies someone who has the possibility of depression, through a process of classifying the model using the Support Vector Machine and t-Distributed Stochastic Neighbor Embedding methods in the DASS-42 data set. The DASS-42 data set consists of 39,776 data and can be used to classify 3 different phenomena namely, depression, stress and anxiety. The maintenance Support Vector Machine model uses DASS-42 data that has been cleaned, normalized and balanced and uses attributes that have been reduced through the t-Distributed Stochastic Neighbor Embedding dimension reduction process. Training data and test data are divided by the ratio of 80:20. Based on the test results, the application of the Support Vector Machine (SVM) and t-Distributed Stochastic Neighbor Embedding (t-SNE) methods for depression classification on DASS-

42 data shows better performance with the best accuracy of 100% on data before balancing and 91, 71% on data after balancing.

Keywords : *Support Vector Machine, t-Distributed Stochastic Neighbor Embedding, Classification, Depression*

1 PENDAHULUAN

Depresi merupakan salah satu gangguan kesehatan mental. Kondisi ini berkaitan dengan hilangnya minat dan keinginan untuk beraktivitas, berpikir negatif, sedih yang berkepanjangan, kesulitan bersosial dan dapat mengarah ke kasus bunuh diri dalam beberapa kasus [1]. Sekitar 300 juta jiwa atau 3.76% populasi di dunia dari segala usia dan komunitas menderita depresi [2]. World Health Organization (WHO) memprediksi bahwa depresi akan menjadi penyebab kematian yang paling berdampak pada masa 15 tahun ke depan. Depresi dapat mempengaruhi kualitas hidup secara signifikan dan menyebabkan masalah baru seperti penurunan prestasi, kehilangan ide kreatif serta kurangnya produktivitas di tempat kerja yang tentunya mempersulit seseorang untuk menikmati kegiatan yang biasanya menyenangkan ([3]–[5]).

Beberapa penelitian menggunakan metode klasifikasi untuk mengidentifikasi seseorang yang memiliki kemungkinan gejala depresi[6]–[8]. Salah satu metode klasifikasi yang sering digunakan dalam klasifikasi depresi adalah Support Vector Machine (SVM) [8], [9]. Metode ini digunakan untuk menemukan pembatas (hyperplane) yang optimal diantara set data dari berbagai kelas yang berbeda di dalam ruang vektor. Kelebihan yang dimiliki oleh metode SVM adalah tingkat akurasi yang tinggi dan tidak sensitif terhadap noise di dalam data. Namun metode ini tidak luput dari kekurangan[10], [11]. Jika set data memiliki ukuran yang besar maka waktu yang dibutuhkan untuk pelatihan model menjadi bertambah lama, serta informasi mengenai probabilitas setiap kelas tidak dijelaskan [12]. Beberapa penelitian menunjukkan bahwa metode SVM terbukti cocok digunakan untuk mengklasifikasi berbagai set data dengan mempertahankan tingkat akurasi yang tinggi [13]–[16].

Penggunaan metode t-Distributed Stochastic Neighbor Embedding (t-SNE) untuk mereduksi dimensi data dan mengekstraksi fitur dilakukan dalam penelitian ini. Harapannya, metode t-SNE dapat menyajikan visualisasi data dengan baik. Metode t-SNE merupakan teknik yang digunakan pada data dengan dimensi tinggi yang direduksi ke data yang dimensi yang lebih rendah tanpa mengubah struktur dari data [17]. Cara kerja metode t-SNE adalah dengan mengukur jarak antar titik pada data dalam dimensi tinggi dan menyajikannya dalam dimensi yang lebih rendah. Metode t-SNE juga dapat digunakan untuk memudahkan visualisasi dan eksplorasi karakteristik dari set data [18]. Pemilihan metode t-SNE sebagai preprocessing terbukti cocok digunakan pada data dengan ukuran besar dan membantu visualisasi yang baik tanpa mengubah struktur data[17], [19]–[21].

Set data yang digunakan dalam penelitian ini adalah DASS-42 (Depression Anxiety Stress Scale) yang diperoleh dari Open Psychometrics pada Kaggle. DASS-42 merupakan skala yang digunakan untuk mengukur tingkat depresi, kecemasan dan stres pada individu [22]. Skala ini terdiri dari 42 pertanyaan yang mengelompokkan individu ke dalam beberapa tingkat instrumen yaitu, normal, ringan, sedang, berat dan sangat berat. Fondasi teori dari skala ini

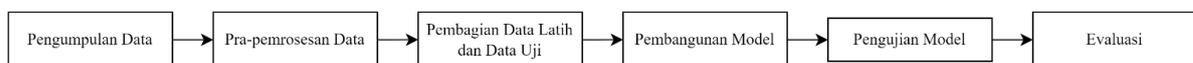
didasarkan pada model tripartite kecemasan dan depresi yang di ajukan oleh Clark dkk. pada tahun 1991 [1]. DASS terdiri dari tiga skala diantaranya, depresi (menilai keadaan mood dysphoria), kecemasan (menilai keadaan gairah) dan stres (menilai labilitas emosi negatif terhadap penyebab dan ketegangan umum) [23]. Terdapat dua versi dari skala ini, ada yang versi 42-item dan versi 21-item. Walaupun Data DASS-42 dapat mengetahui tingkat depresi, kecemasan dan stress individu, penelitian ini berfokus pada tingkat depresi berdasarkan jawaban yang telah diberikan. Data DASS-42 dikoleksi pada tahun 2017 – 2019 terdiri dari 39.775 responden dengan 172 fitur.

Dengan seluruh keunggulan yang telah disebutkan sebelumnya, penelitian ini mengimplementasikan metode SVM dan t-SNE untuk mengklasifikasi depresi pada data DASS-42. Harapannya atribut data DASS-42 dapat diekstrak dengan memanfaatkan visualisasi reduksi dimensi t-SNE dan dapat menghasilkan keakuratan yang baik untuk klasifikasi depresi menggunakan metode SVM. Sehingga apabila seseorang menemukan gejala depresi setelah melakukan deteksi dini menggunakan instrumen DASS-42, dapat direkomendasikan untuk menemui psikolog atau dokter spesialis kejiwaan lebih cepat.

2 METODE PENELITIAN

2.1 GARIS BESAR PENYELESAIAN MASALAH

Implementasi dari metode SVM dan t-SNE untuk klasifikasi depresi pada data DASS-42 memiliki beberapa proses yang dilakukan untuk menyelesaikan masalah. Gambar 1 menjelaskan tahapan yang dilakukan dalam penelitian ini.



Gambar 1. Garis Besar Penyelesaian Masalah

2.2 PENGUMPULAN DATA

Set data yang digunakan berasal dari Open Psychometrics. Data DASS-42 dikoleksi oleh Open Psychometrics pada tahun 2017 – 2019 yang terdiri dari 39.775 responden dengan 172 atribut dan terbagi menjadi 5 kelas depresi yaitu normal, ringan, sedang, berat, dan sangat berat.

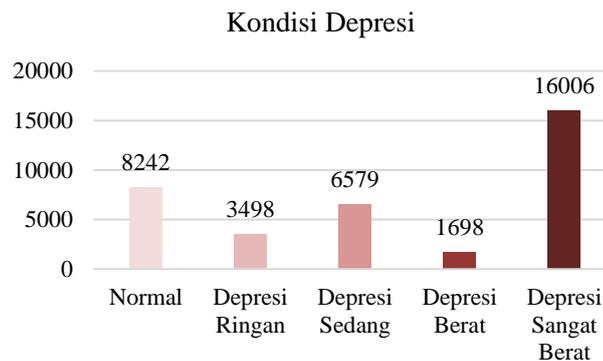
2.3 PRA-PEMROSESAN

Sebelum melakukan pembagian data latih dan data uji, dalam penelitian ini dilakukan pra-pemrosesan data terlebih dahulu.

a. Reduksi Atribut, Pembersihan Data, Pengisian Nilai dan Pemberian Label

Data DASS-42 dalam penelitian ini memiliki beberapa atribut yang tidak berkaitan terhadap perhitungan skor untuk menentukan tingkat depresi seseorang [22]. Selain itu masih ada beberapa atribut yang tidak berkaitan dengan depresi. Sehingga atribut dari Data DASS-42 langsung direduksi dari 172 atribut menjadi 63 atribut. Setelah memilih atribut yang berpengaruh terhadap depresi, tahapan selanjutnya adalah melakukan pembersihan data dan pengisian nilai. Pada tahap ini, setiap atribut yang tidak terdistribusi normal dibersihkan dan

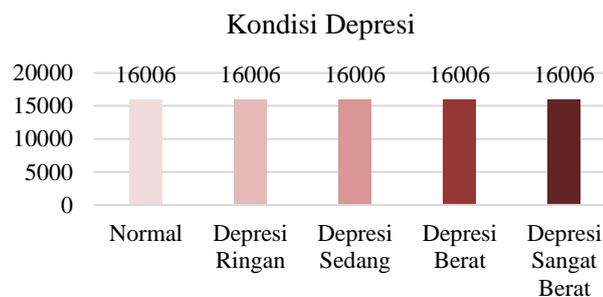
atribut yang memiliki banyak nilai yang hilang atau kosong diisi menggunakan nilai mean dengan pembulatan ke atas. Sehingga jumlah data DASS-42 sebelum dinormalisasi berubah dari 39.775 data menjadi 36.023 data. Setelah melakukan pembersihan data dan pengisian nilai, maka dilakukan tahap pemberian label untuk melengkapi set data agar dapat digunakan dalam penelitian ini. Pemberian label dilakukan sesuai dengan manual pembagian skor DASS-42 [22], [24], [25]. Setelah melakukan langkah pemberian label, maka sebaran kondisi depresi pada data DASS-42 dapat dilihat pada Gambar 2



Gambar 2. Sebaran Kondisi Depresi Data DASS-42

b. Balancing Data

Pada Gambar 2 terdapat beberapa kelas kondisi yang tidak tersebar secara merata sehingga kelas-kelas perlu diseimbangkan agar model yang digunakan dapat mengklasifikasi kondisi depresi secara tepat. Metode yang digunakan untuk melakukan balancing data adalah SMOTE. Sehingga setelah dilakukan balancing data, sebaran kondisi depresi pada Data DASS-42 dapat dilihat pada Gambar 3.



Gambar 3. Sebaran Kondisi Depresi Data DASS-42 setelah Balancing

c. Transformasi Data

Langkah pra-pemrosesan selanjutnya adalah transformasi data. Pada tahap ini dilakukan normalisasi data yang sudah dibersihkan dan diseleksi. Hal ini dilakukan karena terdapat perbedaan nilai yang besar antara satu atribut dengan data lainnya. Langkah normalisasi dilakukan dengan menggunakan Normalisasi Min-Max. Sehingga hasil akhir dari proses ini memiliki atribut yang memiliki nilai 0 – 1.

d. *Proses t-SNE*

Setelah melakukan proses normalisasi, set data divisualisasi menjadi 2 dimensi menggunakan metode t-SNE. Proses ini dilakukan dengan memanfaatkan metode reduksi dimensi t-SNE dengan perplexity = 40 [26]. Langkah yang dilakukan dalam proses t-SNE pada penelitian ini diantaranya:

1. Menghitung jarak antar data
2. Menghitung nilai variansi yang baru menggunakan binary search, probabilitas data pada dimensi asli, dan nilai perplexity yang baru
3. Mengulang proses sebelumnya hingga menemukan variansi paling baru
4. Menghitung probabilitas data pada dimensi baru
5. Menghitung nilai similarity menggunakan persamaan Gradient Descent
6. Mengulangi proses diatas sebanyak 300 iterasi

Dengan melakukan langkah dan perhitungan yang sama, maka diperoleh visualisasi reduksi dimensi t-SNE dengan memanfaatkan scatterplot dari library scikit-learn. Pada penelitian ini, dilakukan 8 skenario fitur data DASS-42 yang berbeda dengan tujuan untuk menghasilkan visualisasi yang terbaik. Kelompok atribut dalam skenario reduksi dimensi t-SNE dibagi menjadi 4 kelompok, yaitu jawaban dari seluruh pertanyaan DASS-42, jawaban dari pertanyaan DASS-42 subskala depresi, jawaban dari TIPI, dan jawaban terkait latar belakang responden. Tabel 1 memperlihatkan visualisasi reduksi dimensi t-SNE.

2.4 PEMBAGIAN DATA LATIH DAN DATA UJI

Setelah melalui pra-pemrosesan, set data DASS-42 dibagi menjadi data latih dan data uji. Data latih digunakan untuk membangun model klasifikasi dengan menggunakan metode SVM, sedangkan data uji digunakan untuk mengevaluasi model. Set data DASS-42 yang berjumlah 80.030 data dibagi dengan komposisi perbandingan jumlah data latih dan data uji adalah 80:20 [8]. Sehingga 64.024 data digunakan menjadi data latih dan 16.006 data digunakan sebagai data uji.

2.5 PELATIHAN MODEL SVM

Set data latih selanjutnya melakukan proses pelatihan menggunakan metode Support Vector Machine One-Against-All dengan memanfaatkan kernel Polinomial. Berikut adalah proses pelatihan model SVM OAA menggunakan kernel Polinomial.

1. Menghitung kernel polinomial
Pada proses ini ditetapkan bahwa $c = 1$ [27] dan $\text{degree} = 2$ [28].
2. Menghitung matiks *hessian*
Inisialisasi awal $\lambda = 0,5$ [29]
3. Menghitung nilai *epsilon*
Inisialisasi awal $\alpha = 0$ [30]

4. Menghitung nilai *delta alpha*
Inisialisasi awal $\gamma = 0,00001$ dan $C=1$ [29].
5. Menghitung nilai *alpha*
6. Menghitung *update* nilai *epsilon*
7. Menghitung *update* nilai *delta alpha*
8. Menghitung *update* nilai *alpha*
9. Menghitung kernel positif dan negatif
Kernel positif didapatkan melalui pencarian nilai positif terbesar dan kernel negatif didapatkan pada nilai negatif terbesar pada matriks hessian.
10. Menghitung bobot positif dan negatif
11. Menghitung nilai bias

2.6 PENGUJIAN MODEL SVM

Pengujian model SVM pada metode Sequential Training dilakukan dengan melakukan perhitungan fungsi diskriminan $f(x)$ dan menggunakan data uji. Jika hasil dari $f(x)$ positif maka diklasifikasikan sebagai kelas positif, dan begitu juga sebaliknya.

2.7 EVALUASI

Untuk mengukur akurasi dan performa dari model yang telah dibangun, perlu dilakukan evaluasi model. Penelitian ini menggunakan metode evaluasi bernama k-fold cross validation. Evaluasi hasil dari penelitian ini dilakukan dengan menghitung rata-rata akurasi dari sebanyak k-fold. Untuk setiap iterasi yang berbeda berdasarkan pasangan atribut dan visualisasi t-SNE dibuat sebanyak 10 lipatan. Nilai akurasi dapat dihitung dengan menjumlah data uji yang diprediksi benar dan dibagi dengan jumlah seluruh data uji. Kemudian hasil dari rata-rata akurasi dari setiap k dibandingkan.

3 HASIL PENELITIAN

3.1 SKENARIO PENGUJIAN

Skenario pengujian dilakukan untuk menguji model Support Vector Machine terhadap data DASS-42. Dalam penelitian ini terdapat 3 skenario pengujian. Skenario yang dilakukan diantara lain adalah membandingkan pasangan terbaik berdasarkan hasil visualisasi reduksi dimensi t-SNE, membandingkan set data DASS-42 sebelum dan sesudah balancing menggunakan SMOTE, dan memvalidasi performa pengujian berdasarkan rata-rata akurasi sebanyak k dari k-fold cross validation. Ketiga skenario pengujian dilakukan dengan menggunakan Support Vector Machine One-Against-All kernel Polinomial dengan nilai degree = 2 dan $c = 1$ [27], [28].

3.2 SKENARIO PENGUJIAN 1: PERBANDINGAN PASANGAN ATRIBUT

Skenario pengujian pertama dilakukan untuk membandingkan pasangan atribut berdasarkan skenario visualisasi reduksi dimensi t-SNE. Dalam skenario pengujian 1, data yang sudah melalui pra-pemrosesan tanpa balancing dilakukan perbandingan pengujian menggunakan 8 pasangan atribut data yang berbeda seperti pada Tabel 1. Tabel 2 memperlihatkan hasil dari skenario pengujian 1.

Tabel 2. Hasil Pengujian Skenario 1.

<i>Skenario</i>	<i>Pasangan Atribut</i>	<i>Akurasi</i>
1	Q1A - Q42A, TIPI1 - TIPI10, <i>Education, Urban, Gender, Age, Hand, Religion, Orientation, Race, Voted, Married, Familysize</i>	99,49%
2	Q1A - Q42A, TIPI1 - TIPI10	99,54%
3	Q1A - Q42A, <i>Education, Urban, Gender, Age, Hand, Religion, Orientation, Race, Voted, Married, Familysize</i>	99,04%
4	Q1A - Q42A	98,79%
5	Q3A, Q5A, Q10A, Q13A, Q16A, Q17A, Q21A, Q24A, Q26A, Q31A, Q34A, Q37A, Q38A, Q42A, TIPI1 - TIPI10, <i>Education, Urban, Gender, Age, Hand, Religion, Orientation, Race, Voted, Married, Familysize</i>	99,92%
6	Q3A, Q5A, Q10A, Q13A, Q16A, Q17A, Q21A, Q24A, Q26A, Q31A, Q34A, Q37A, Q38A, Q42A, TIPI1 - TIPI10	99,93%
7	Q3A, Q5A, Q10A, Q13A, Q16A, Q17A, Q21A, Q24A, Q26A, Q31A, Q34A, Q37A, Q38A, Q42A, <i>Education, Urban, Gender, Age, Hand, Religion, Orientation, Race, Voted, Married, Familysize</i>	99,96%
8	Q3A, Q5A, Q10A, Q13A, Q16A, Q17A, Q21A, Q24A, Q26A, Q31A, Q34A, Q37A, Q38A, Q42A	100%

Pasangan atribut pada Skenario Pengujian 1 dipasangkan berdasarkan kelompok jawaban dari set data DASS-42. Kelompok jawaban dari set data DASS-42 diantaranya adalah jawaban dari seluruh pertanyaan DASS-42, jawaban dari pasangan pertanyaan depresi, jawaban dari TIPI, serta jawaban mengenai latar belakang responden. Pada Tabel 4.1 hasil pengujian menunjukkan bahwa skenario ke-8 memperoleh akurasi terbaik. Pasangan atribut skenario ke-8 memiliki skor akurasi sebesar 100%. Hal ini bisa terjadi karena atribut yang digunakan dalam prosesnya adalah jawaban dari pertanyaan yang berkaitan dengan fenomena depresi, serta visualisasi yang dihasilkan melalui reduksi dimensi t-SNE menggambarkan sebaran kondisi kelas depresi pada data DASS-42 yang lebih terkelompok dan memiliki noise yang lebih sedikit.

3.3 SKENARIO PENGUJIAN 2: PERBANDINGAN DATA SEBELUM DAN SESUDAH BALANCING

Skenario pengujian kedua dilakukan untuk membandingkan set data DASS-42 yang digunakan dalam pemodelan SVM OAA. Dalam skenario ini, set data DASS-42 yang belum diseimbangkan dibandingkan dengan yang sudah diseimbangkan menggunakan metode SMOTE. Tabel 3 memperlihatkan hasil dari skenario pengujian 2.

Tabel 3. Hasil Pengujian Skenario 2.

<i>Skenario</i>	<i>Akurasi</i>	
	<i>Data dengan Balancing</i>	<i>Data tanpa Balancing</i>
1	86,51%	99,49%
2	86,51%	99,54%
3	86,73%	99,04%
4	87,04%	98,79%
5	86,07%	99,92%
6	85,53%	99,93%
7	86,10%	99,96%
8	91,71%	100%

Hasil pengujian yang telah dilakukan pada Skenario 2 menunjukkan bahwa proses balancing data yang dilakukan pada set data DASS-42 memperoleh performa yang tidak lebih baik dari yang sebelum balancing. Akurasi yang dihasilkan oleh data sebelum dan sesudah balancing memiliki perbedaan sebesar 9-13%. Hal ini dapat disebabkan oleh overfitting yang terjadi pada label kelas yang dilakukan balancing [31]. Meskipun demikian, hasil dari pengujian skenario ke-2 tetap membuktikan bahwa reduksi dimensi t-SNE menghasilkan akurasi terbaik pada data balancing maupun tanpa balancing.

3.4 SKENARIO PENGUJIAN 3: VALIDASI PERFORMA

Skenario pengujian ketiga dilakukan untuk memvalidasi hasil dari skenario pengujian pertama dan kedua. Validasi dilakukan dengan membandingkan hasil akurasi dari perhitungan pada pengujian 1 dan 2 berdasarkan k-fold yang diinisialisasi. Penelitian ini menggunakan 10-fold untuk memvalidasi hasil skenario pengujian 1 dan 2. Tabel 4 memperlihatkan hasil dari skenario pengujian 3.

Tabel 4. Hasil Pengujian Skenario 3.

K	Skenario	Akurasi	
		Data dengan Balancing	Data tanpa Balancing
10	1	86,51%	99,49%
	2	86,51%	99,54%
	3	86,73%	99,04%
	4	87,04%	98,79%
	5	86,07%	99,92%
	6	85,53%	99,93%
	7	86,10%	99,96%
	8	91,71%	100%

Hasil pengujian yang telah dilakukan pada Tabel 4 menunjukkan bahwa validasi 10-Fold Cross Validation pada skenario ke-8 reduksi dimensi t-SNE masih memperoleh rata-rata akurasi terbaik. Akurasi yang dihasilkan dari validasi tidak jauh berbeda dengan akurasi dari pengujian skenario ke-2.

4 HASIL DAN PEMBAHASAN

Berdasarkan hasil pengujian skenario 1 sampai dengan skenario 3 menggunakan set data DASS-42 menggunakan metode Support Vector Machine One-Against-All kernel Polinomial, berikut adalah beberapa hasil analisa yang diperoleh

Pada pengujian skenario 1, akurasi terbaik dihasilkan oleh pasangan atribut skenario ke-8 reduksi dimensi t-SNE. Dengan jumlah atribut sebanyak 14 dan nilai akurasi yang dihasilkan sebesar 100%. Pasangan atribut pada skenario ke-8 merupakan jawaban dari pertanyaan yang berhubungan dengan fenomena depresi. Pada pengujian skenario 2, set data DASS-42 sesudah diseimbangkan tidak memperoleh hasil akurasi yang lebih tinggi dari sebelum diseimbangkan. Akurasi terbaik yang dihasilkan oleh set data DASS-42 setelah balancing adalah 91,71% pada skenario ke-8. Sedangkan nilai akurasi terendah berada pada nilai 85,53%. Meskipun demikian, pengujian skenario 2 menunjukkan bahwa reduksi dimensi t-SNE dengan skenario ke-8 merupakan pasangan atribut dengan akurasi terbaik.

Selanjutnya pada pengujian skenario ke-3, data dengan balancing SMOTE dan tanpa balancing dilakukan validasi dengan menggunakan 10-fold cross validation. Hasil rata-rata akurasi terbaik yang dihasilkan tidak berbeda jauh yaitu sebesar 91,71% untuk setelah balancing dan 100% untuk sebelum balancing. Metode Support Vector Machine dapat digunakan untuk mengklasifikasi depresi pada data DASS-42. Performa terbaik dihasilkan dengan menggunakan pasangan atribut skenario ke-8 reduksi dimensi t-SNE, dan menggunakan data yang belum diseimbangkan. Akurasi terbaik yang dihasilkan sebesar 100% telah divalidasi menggunakan 10-Fold Cross Validation.

5 KESIMPULAN

Berdasarkan penelitian mengenai Implementasi Metode Support Vector Machine (SVM) dan t-Distributed Stochastic Neighbor Embedding (t-SNE) untuk Klasifikasi Depresi dapat disimpulkan bahwa implementasi metode Support Vector Machine dan t-Distributed Stochastic Neighbor Embedding (t-SNE) dalam mengklasifikasi kondisi depresi pada data Depression Anxiety Stress Scales (DASS-42) menunjukkan performa yang lebih baik. Selain itu, klasifikasi depresi pada data Depression Anxiety Stress Scales (DASS-42) menunjukkan performa yang lebih baik dengan menggunakan set data yang belum dibalancing. Untuk penelitian selanjutnya, metode dapat diimplementasikan pada kasus data set lain untuk mengetahui lebih dalam dan lebih detail mengenai performa yang dihasilkan, berikut kelebihan dan kekurangannya.

DAFTAR PUSTAKA

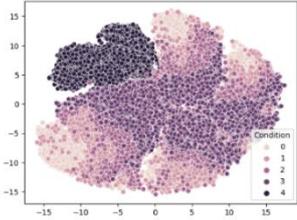
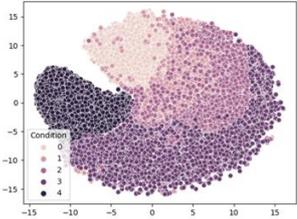
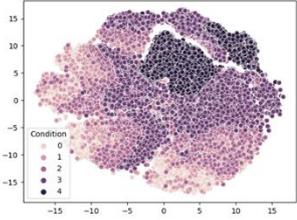
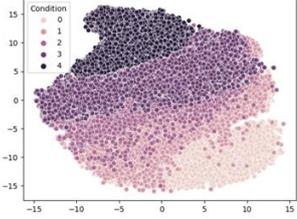
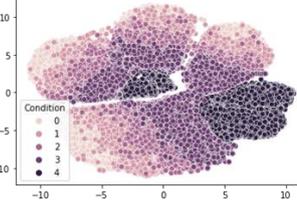
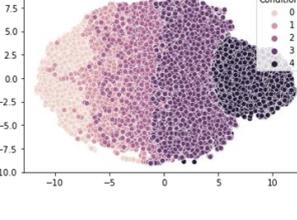
- [1] R. C. B. Vignola and A. M. Tucci, "Adaptation and Validation of the Depression, Anxiety and Stress Scale (DASS) to Brazilian Portuguese," *J Affect Disord*, vol. 155, pp. 104–109, 2014, doi: <https://doi.org/10.1016/j.jad.2013.10.031>.
- [2] J. Balbuena, S. Almeyda, J. Mendoza, and J. A. Pow-Sang, "Depression Detection Using Audio-Visual Data and Artificial Intelligence: A Systematic Mapping Study," in *Proceedings of Fifth International Congress on Information and Communication Technology*, S. and D. N. and J. A. Yang Xin-She and Sherratt, Ed., Singapore: Springer Singapore, 2021, pp. 296–306.
- [3] H. Dibeklioglu, Z. Hammal, and J. F. Cohn, "Dynamic Multimodal Measurement of Depression Severity Using Deep Autoencoding," *IEEE J Biomed Health Inform*, vol. 22, no. 2, pp. 525–536, 2018, doi: [10.1109/JBHI.2017.2676878](https://doi.org/10.1109/JBHI.2017.2676878).
- [4] K. Dianovinina, "Depresi pada Remaja: Gejala dan Permasalahannya," *Jurnal Psikogenesis*, vol. 6, no. 1, 2018, doi: <https://doi.org/10.24854/jps.v6i1.634>.
- [5] P. E. Greenberg, A. A. Fournier, T. Sisitsky, C. T. Pike, and R. C. Kessler, "The economic burden of adults with major depressive disorder in the United States (2005 and 2010)," *Journal of Clinical Psychiatry*, vol. 76, no. 2, pp. 155–162, Feb. 2015, doi: [10.4088/JCP.14m09298](https://doi.org/10.4088/JCP.14m09298).
- [6] A. Budiman, J. C. Young, and A. Suryadibrata, "Implementasi Algoritma Naive Bayes untuk Klasifikasi Konten Twitter dengan Indikasi Depresi," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 6, no. 2, 2021.

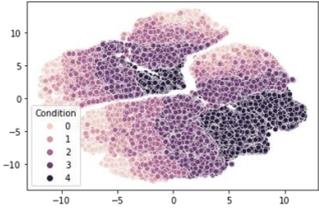
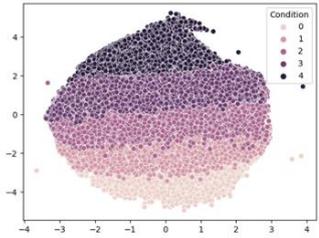
- [7] N. Aini, "Penerapan Jaringan Syaraf Tiruan dengan Metode Radial Basis Function (RBF) untuk Klasifikasi Gangguan Depresi," Universitas Islam Negeri Sultan Syarif Kasim Riau, 2021.
- [8] S. Aprilla, M. T. Furqon, and M. A. Fauzi, "Klasifikasi Penyakit Skizofrenia dan Episode Depresi Pada Gangguan Kejiwaan Dengan Menggunakan Metode Support Vector Machine (SVM)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, vol. 2548, p. 964X, 2018.
- [9] K. S. Srinath, K. Kiran, S. Pranavi, M. Amrutha, P. D. Shenoy, and K. R. Venugopal, "Prediction of Depression, Anxiety and Stress Levels Using Dass-42," in 2022 IEEE 7th International conference for Convergence in Technology, I2CT 2022, Institute of Electrical and Electronics Engineers Inc., Apr. 2022. doi: 10.1109/I2CT54291.2022.9824087.
- [10] C. P. Diehl and G. Cauwenberghs, "SVM Incremental Learning, Adaptation and Optimization," in *Proceedings of the International Joint Conference on Neural Networks*, 2003., 2003, pp. 2685–2690 vol.4. doi: 10.1109/IJCNN.2003.1223991.
- [11] R. Cosme and R. Krohling, "Support Vector Machines Applied to Noisy Data Classification Using Differential Evolution with Local Search," Jan. 2011.
- [12] K. Mujib, A. Hidayatno, and T. Prakoso, "Pengenalan Wajah Menggunakan Local Binary Pattern (LBP) Dan Support Vector Machine (SVM)," *Transient: Jurnal Ilmiah Teknik Elektro*, vol. 7, no. 1, pp. 123–130, 2018, doi: <https://doi.org/10.14710/transient.v7i1.123-130>.
- [13] W. Agustina, M. T. Furqon, and B. Rahayudi, "Implementasi Metode Support Vector Machine (SVM) Untuk Klasifikasi Rumah Layak Huni (Studi Kasus: Desa Kidal Kecamatan Tumpang Kabupaten Malang)," vol. 2, no. 10, pp. 3366–3372, 2018, Accessed: Mar. 02, 2023. [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2615>
- [14] H. Peiqing, "Multidimensional State Data Reduction and Evaluation of College Students' Mental Health Based on SVM," *Journal of Mathematics*, vol. 2022, 2022, doi: 10.1155/2022/4961203.
- [15] N. A. Utami, W. Maharani, and I. Atastina, "Personality Classification of Facebook Users According to Big Five Personality Using SVM (Support Vector Machine) Method," *Procedia Comput Sci*, vol. 179, pp. 177–184, Jan. 2021, doi: 10.1016/J.PROCS.2020.12.023.
- [16] Y. Li and F. Fan, "Classification of Schizophrenia and Depression by EEG with ANNs*," in 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, 2005, pp. 2679–2682. doi: 10.1109/IEMBS.2005.1617022.
- [17] B. Melit Devassy and S. George, "Dimensionality Reduction and Visualisation of Hyperspectral Ink Data using t-SNE," *Forensic Sci Int*, vol. 311, p. 110194, Jun. 2020, doi: 10.1016/J.FORSCIINT.2020.110194.
- [18] F. Anowar, S. Sadaoui, and B. Selim, "Conceptual and Empirical Comparison of Dimensionality Reduction Algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE)," *Comput Sci Rev*, vol. 40, p. 100378, May 2021, doi: 10.1016/J.COSREV.2021.100378.
- [19] A. Russo and A. Borrás, "Comparison of Dimension Reduction Techniques applied to the Analysis of Airborne Radionuclide Activity Concentration," *J Environ Radioact*, vol. 244–245, p. 106813, Apr. 2022, doi: 10.1016/J.JENVRAD.2022.106813.
- [20] C. Xiao, F. Khayatian, and G. Dall'O', "Unsupervised Learning for Feature Projection: Extracting Patterns from Multidimensional Building Measurements," *Energy Build*, vol. 224, p. 110228, Oct. 2020, doi: 10.1016/J.ENBUILD.2020.110228.

- [21] A. K. Abbas, A. I. Khalil, and S. A. Abdulkader, "Social Touch Recognition Based on Support Vector Machine and T-Distributed Stochastic Neighbour Embedding as Pre-processing," *IOP Conf Ser Mater Sci Eng*, vol. 1076, no. 1, p. 012042, Feb. 2021, doi: 10.1088/1757-899x/1076/1/012042.
- [22] S. H. Lovibond and P. F. Lovibond, *Manual for the Depression, Anxiety and Stress Scales (DASS)*, vol. 2. 1995.
- [23] P. J. Norton, "Depression Anxiety and Stress Scales (DASS-21): Psychometric Analysis Across Four Racial Groups," *Anxiety Stress Coping*, vol. 20, no. 3, pp. 253–265, Sep. 2007, doi: 10.1080/10615800701309279.
- [24] M. M. Antony and R. P. Swinson, "Psychometric Properties of the 42-item and 21-item Versions of the Depression Anxiety Stress Scales in Clinical Groups and A Community Sample," *Psychol Assess*, Jun. 1998, doi: 10.1037/1040-3590.10.2.176.
- [25] A. Afzali, A. Delavar, A. Borjali, and S.-M. Mirzamani, "Psychometric Properties of DASS-42 as Assessed in a Sample of Kermanshah High School students.," *Journal of Research in Behavioural Sciences*, vol. 5, pp. 81–92, Jan. 2007.
- [26] L. van der Maaten and G. Hinton, "Visualizing Data using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. 86, pp. 2579–2605, 2008, [Online]. Available: <http://jmlr.org/papers/v9/vandermaaten08a.html>
- [27] H. Al Azies, D. Trishnanti, and E. Mustikawati, "Comparison of Kernel Support Vector Machine (SVM) in Classification of Human Development Index (HDI)," in *IPTEK Journal of Proceedings Series*, 2019. doi: <http://dx.doi.org/10.12962/j23546026.y2019i6.6394>.
- [28] S. Ali and K. Smith-Miles, "On Optimal Degree Selection for Polynomial Kernel with Support Vector Machines: Theoretical and Empirical Investigations," *International Journal of Knowledge-Based and Intelligent Engineering Systems*, vol. 11, no. 1, pp. 1–18, 2007, doi: 10.3233/KES-2007-11101.
- [29] R. A. Wijayanti, M. T. Furqon, and S. Adinugroho, "Penerapan Algoritme Support Vector Machine Terhadap Klasifikasi Tingkat Risiko Pasien Gagal Ginjal," vol. 2, no. 10, pp. 3500–3507, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [30] S. Vijayakumar and S. Wu, "Sequential Support Vector Classifiers and Regression," Mar. 1999.
- [31] M. Decuyper, M. Stockhoff, S. Vandenberghe, al -, and X. Ying, "An Overview of Overfitting and its Solutions," *J Phys Conf Ser*, vol. 1168, no. 2, p. 022022, Feb. 2019, doi: 10.1088/1742-6596/1168/2/022022.

LAMPIRAN

Tabel 1. Hasil Visualisasi Rediksi Dimensi t-SNE

Skenario	Atribut	Visualisasi	Keterangan	Waktu Pemrosesan
1	Q1A - Q42A, TIPI1 - TIPI10, Education, Urban, Gender, Age, Hand, Religion, Orientation, Race, Voted, Married, Familysize		Pasangan atribut dari skenario 1 terdiri dari seluruh atribut dari set data DASS-42.	Waktu: 8 menit 54 detik KL Divergence: 4,464113
2	Q1A - Q42A, TIPI1 - TIPI10		Pasangan atribut dari skenario 2 terdiri dari seluruh atribut dari set data DASS-42 tanpa latar belakang responden.	Waktu: 11 menit 19 detik KL Divergence: 4,735845
3	Q1A - Q42A, Education, Urban, Gender, Age, Hand, Religion, Orientation, Race, Voted, Married, Familysize		Pasangan atribut dari skenario 3 merupakan seluruh atribut dari set data DASS-42 tanpa jawaban TIPI atau pertanyaan kepribadian.	Waktu: 11 menit 1 detik KL Divergence: 4,355434
4	Q1A - Q42A		Pasangan atribut dari skenario 4 hanya terdiri dari jawaban dari 42 pertanyaan DASS-42	Waktu: 9 menit 10 detik KL Divergence: 4,652343
5	Q3A, Q5A, Q10A, Q13A, Q16A, Q17A, Q21A, Q24A, Q26A, Q31A, Q34A, Q37A, Q38A, Q42A, TIPI1 - TIPI10, Education, Urban, Gender, Age, Hand, Religion, Orientation, Race, Voted, Married, Familysize		Pasangan atribut dari skenario 5 terdiri dari jawaban pertanyaan yang berkorelasi dengan Depresi, jawaban TIPI dan latar belakang responden	Waktu: 9 menit 17 detik KL Divergence: 4,525468
6	Q3A, Q5A, Q10A, Q13A, Q16A, Q17A, Q21A, Q24A, Q26A, Q31A, Q34A, Q37A, Q38A, Q42A, TIPI1 - TIPI10		Pasangan atribut dari skenario 6 terdiri dari jawaban pertanyaan yang berkorelasi dengan depresi, dan jawaban TIPI atau pertanyaan kepribadian	Waktu: 9 menit 20 detik KL Divergence: 4,878942

Skenario	Atribut	Visualisasi	Keterangan	Waktu Pemrosesan
7	Q3A, Q5A, Q10A, Q13A, Q16A, Q17A, Q21A, Q24A, Q26A, Q31A, Q34A, Q37A, Q38A, Q42A, <i>Education, Urban, Gender, Age, Hand, Religion, Orientation, Race, Voted, Married, Familysize</i>		Pasangan atribut dari skenario 7 terdiri dari jawaban pertanyaan yang berkorelasi dengan depresi dan latar belakang responden	Waktu: 9 menit 14 detik KL Divergence: 4,281454
8	Q3A, Q5A, Q10A, Q13A, Q16A, Q17A, Q21A, Q24A, Q26A, Q31A, Q34A, Q37A, Q38A, Q42A		Pasangan atribut dari skenario 8 hanya terdiri dari jawaban pertanyaan yang berkorelasi dengan depresi.	Waktu: 9 menit 12 detik KL Divergence: 5,556355