

IMPLEMENTASI MACHINE LEARNING UNTUK MEMPREDIKSI TINGKAT STRES BERDASARKAN GAYA HIDUP

Widyasmara Afif Nur Rahman; Endang Wahyu Pamungkas

Program Studi Teknik Informatika, Fakultas Komunikasi dan Informatika,
Universitas Muhammadiyah Surakarta

Abstrak

Stress adalah isu utama saat ini, terutama pada generasi muda. Masalah mental ini dapat mengarah pada berbagai masalah seperti depresi, serangan jantung, stress dan lebih buruk lagi bunuh diri. Tujuan dari penelitian ini adalah untuk memprediksi tingkat stress berdasarkan gaya hidup seseorang. Penulis akan melakukan analisis faktor faktor gaya hidup serta mencari korelasi antara faktor gaya hidup dan tingkat stress. Dataset untuk penelitian ini diambil dari Kaggle dan terdiri dari 15,977 data. Algoritma klasifikasi yang digunakan yaitu Support Vector Machine (SVM) dan Random Forest, Akurasi akan digunakan sebagai parameter kinerja. Selanjutnya akan ditentukan algoritma yang lebih baik yang akan digunakan untuk di deploy menggunakan streamlit. Kemudian aplikasi akan diuji dengan metode System Usability Scale. Pembuatan model akan dilakukan dalam tiga percobaan, tanpa pengubahan data, Pengelompokan Kategori, dan Pengelompokan Kategori dan Penanganan data Imbalance. Dari percobaan yang dilakukan, diperoleh akurasi tertinggi sebesar 0.7667 menggunakan model dengan Pengelompokan Kategori dan ROS. Model yang didapat kemudian dibuat menjadi website sederhana dengan streamlit, website yang dibuat memiliki tiga bagian halaman yaitu halaman landing page, halaman pengisian data, halaman hasil prediksi. Sistem yang dibuat diuji dengan System Usability Scale (SUS) dengan 30 responden dan mendapatkan hasil skor rata-rata 76.5 dengan kategori Good and Acceptable dengan grade C.

Kata Kunci: machine learning, stress, klasifikasi, svm, random forest, streamlit

Abstract

Stress is a major issue today, especially in the younger generation. This mental problem can lead to various problems such as depression, heart attacks, stress and worse, suicide. The aim of this research is to predict stress levels based on a person's lifestyle. The author will conduct a factor analysis of lifestyle factors and look for correlations between lifestyle factors and stress levels. The dataset for this research was taken from Kaggle and consists of 15,977 data. The classification algorithms used are Support Vector Machine (SVM) and Random Forest. Accuracy will be used as a performance parameter. Next, a better algorithm will be determined that will be used to deploy using Streamlit. Then the application will be tested using the System Usability Scale method. Model building will be carried out in three experiments, without data modification, Category Grouping, and Category Grouping and Handling Imbalance data. From the experiments carried out, the highest accuracy was obtained at 0.7667 using a model with Category Grouping and ROS. The model obtained was then made into a simple website with streamlit. The website created had three page sections, namely a landing page, a data entry page, and a prediction results page. The system created was tested using the System Usability Scale (SUS) with 30 respondents and obtained an average score of 76.5 in the Good and Acceptable category with grade C.

Keywords: machine learning, stress, classification, svm, random forest, streamlit

1. PENDAHULUAN

Stres menjadi masalah kesehatan yang sering dialami seseorang dan dapat mempengaruhi kesejahteraan seseorang. Stres adalah reaksi dari dalam diri kita sebagai proses penyesuaian diri dengan situasi yang sedang kita alami (Fadhila et al., 2023). Stres juga diartikan sebagai hasil interaksi dinamis antara individu dengan lingkungannya, yang melibatkan perasaan individu (Ulfa & Rizqi, 2019). Stres menjadi masalah yang masih menghantui masyarakat Indonesia pada saat ini khususnya pada Generasi Z bangsa Indonesia, dan mungkin akan terjadi krisis Kesehatan Mental pada Generasi Z Indonesia, hal ini juga pernah diteliti oleh University College London yang menyatakan bahwa Gen Z lebih rentan terkena gangguan mental, khususnya Depresi.

Berdasarkan Data hasil Riset pada tahun 2018 tentang Kesehatan Dasar (Riskesdas 2018), anggota rumah tangga yang mengidap skizofrenia mengalami peningkatan dari 1,7 per mil pada tahun 2013 menjadi sebesar 7 per mil pada tahun 2018. Dari umur 15 tahun ke atas juga meningkat dari sekitar 12 juta penduduk atau sebesar 6,1% menjadi sekitar 20 juta penduduk atau sebesar 9,8%. Hal ini semakin memburuk akibat pandemi Covid-19, dengan meningkatnya laporan tentang masalah kesehatan jiwa atas dampak dari pandemi tersebut (Kemenkes, 2013, 2018). Sekitar 5,5% remaja dengan usia 10-17 tahun didiagnosa mengalami gangguan mental dalam kurun waktu 12 bulan terakhir, orang yang mengalami gangguan mental biasa disebut dengan Orang Dengan Gangguan Jiwa (ODGJ). Sementara itu, sekitar 34,9% atau sepertiga dari Masyarakat memiliki setidaknya satu masalah kesehatan mental, tergolong Orang Dengan Masalah Kejiwaan (ODMK) (National Adolescent Mental Health Survey, 2022).

Masalah kesehatan mental bisa menjadi sangat berbahaya karena dapat memicu ide untuk bunuh diri. Ada keterkaitan antara bunuh diri dan tingkat stres, di mana semakin tinggi tingkat stres, semakin besar kemungkinan munculnya ide bunuh diri (Lalenoh et al., 2021). Belakangan ini, kasus bunuh diri semakin sering terjadi, dan di kolom komentar media sosial yang membahas berita bunuh diri tersebut, terdapat balasan seperti "Tunggu aku kak, aku sebentar lagi menyusul," yang menunjukkan banyak netizen Indonesia memiliki gangguan mental. Situasi ini bisa saja merugikan terutama bagi bangsa Indonesia, karena dapat menyebabkan hilangnya generasi penerus yang diharapkan mampu mengembangkan negara.

Menurut data yang dibuat oleh Kementerian Dalam Negeri (Kemendagri), jumlah penduduk Indonesia mencapai 277,75 juta jiwa per 31 Desember, dengan mayoritas berumur 10-19 tahun sebanyak 46,2 juta jiwa. Berdasarkan data tersebut, terdapat 2,54 juta remaja yang tergolong ODGJ dan 16,1 juta yang tergolong ODMK. Persentase ini sangat tinggi. Bahkan, survei

lanjutan menunjukkan munculnya kecenderungan bunuh diri. Dari seluruh sampel, dilaporkan 1,4% memiliki ide bunuh diri, 0,5% telah merencanakan bunuh diri, 0,2% mencoba bunuh diri, dan 0,4% pernah melakukan percobaan bunuh diri selama hidupnya. Dari laporan tersebut, 80% remaja yang melaporkan perilaku bunuh diri (memiliki ide, merencanakan, dan mencoba) mengalami masalah gangguan mental.

Untuk membantu mengurangi munculnya tindakan bunuh diri, bisa dilakukan pendeteksian awal sebelum seseorang semakin terpuruk dalam bayang bayang stress. Adapun beberapa penelitian terdahulu tentang pendeteksian tingkat stress contohnya *Mental Stress Detection In University*, *Stress Detection Using Machine Learning*, Deteksi Stres Manusia Melalui Analisis Tidur Dengan Metode Fuzzy (Ahuja & Banga, 2019; Hanchate et al., 2023; Zamzani et al., 2023) . Dan masih banyak lagi penelitian tentang pendeteksian stress dengan menggunakan parameter yang berbeda.

Pada penelitian ini, penulis menggunakan gaya hidup sebagai parameter untuk mengukur tingkat stress, karena gaya hidup mencakup aspek – aspek yang secara langsung mempengaruhi kesejahteraan mental dan fisik seseorang. Seseorang yang memiliki gaya hidup yang buruk cenderung memiliki gangguan mental, pada beberapa penelitian mengindikasikan bahwa semakin tinggi tingkat stress, maka semakin banyak gaya hidup tidak sehat yang dilakukan (Maruyama & Morimoto, 1997; Omasu et al., 2022)

Penelitian ini bertujuan untuk memprediksi tingkat stress seseorang berdasarkan gaya hidup seseorang dengan menggunakan algoritma machine learning dan akan di deploy dalam bentuk web sederhana, yang diharapkan dapat membantu untuk mendeteksi tingkat stress seseorang lebih awal serta memberikan pengarahannya / himbauan untuk menangani tingkat stress berdasarkan data yang dimasukkan oleh user. Website ini ditargetkan dapat digunakan para psikiater untuk mendeteksi tingkat stress pasien lebih awal.

2. METODE

Penelitian ini akan dimulai dari pengumpulan data, setelah data diperoleh maka dilakukan preprocessing data, dilanjutkan dengan pembuatan model machine learning, model tersebut akan dievaluasi untuk menentukan model mana yang akan digunakan, setelah model dipilih maka selanjutnya akan di deploy dengan streamlit, lalu yang terakhir hasil deploy akan dilakukan pengujian untuk melihat seberapa layak website tersebut. Adapun metode penelitian yang lebih jelas ada pada Gambar 1.



Gambar 1. Metode Penelitian

2.1 Pengumpulan Data

Pada penelitian ini menggunakan dataset “Lifestyle and Wellbeing Data” yang diperoleh dari Kaggle.com(Dalat, 2021). Data dari Kaggle terdiri dari 15,977 data dengan 24 fitur, yang ditampilkan pada tabel 1.

Tabel 1. Fitur dari dataset Lifestyle and Wellbeing Data

No	Nama Atribut	Deskripsi
1	Timestamp	Tanggal Data Diperoleh
2	Fruits Veggies	Buah yang dikonsumsi per-hari
3	Daily Stress	Tingkat stress harian seseorang
4	Places Visited	Jumlah tempat baru yang dikunjungi dalam 3 bulan terakhir
5	Core Circle	Jumlah orang yang sangat dekat
6	Supporting Other	Seberapa sering membantu orang lain
7	Social Network	Berapa banyak orang yang berkomunikasi dalam sehari
8	Achievement	Jumlah pencapaian luar biasa yang dapat dibanggakan
9	Donation	Tingkat keserangan dalam menyediakan waktu / uang untuk tujuan baik
10	BMI Range	Kategori BMI
11	To Do Completed	Tingkat Keserangan dalam menyelesaikan daftar mingguan
12	Flow	Mengikuti Alur/ melakukan hal yang sama dengan hari kemarin (jam)
13	Daily Steps	Jumlah langkah yang dilakukan dalam sehari
14	Live Vision	Seberapa jauh pandangan tentang hidup kedepannya
15	Sleep Hours	Lama Tidur
16	Lost Vacation	Jumlah hari yang dihabiskan dalam liburan tahunan
17	Daily Shouting	Tingkat keserangan seseorang dalam berteriak / merajuk pada seseorang
18	Sufficient Income	Kecukupan penghasilan untuk menutup biaya hidup dasar
19	Personal Awards	Jumlah pengakuan yang telah diperoleh dalam hidup
20	Time For Passion	Jumlah waktu yang dihabiskan untuk melakukan sesuatu yang diinginkan
21	Weekly Meditation	Tingkat keserangan seseorang untuk berpikir tentang diri sendiri
22	Age	Umur
23	Gender	Jenis Kelamin
24	Work Life Balance	Hasil Kalkulasi skor kepuasan hidup (Life Satisfactory)

Data yang diperoleh berdasarkan hasil survey yang dilakukan oleh website 360° Living, yang dapat dikunjungi pada <http://www.authentic-happiness.com/> . Survey tersebut bertujuan untuk mengetahui skor kepuasan hidup (Life Satisfactory), Ada 24 atribut yang menjelaskan cara kita menjalani hidup, termasuk kolom kalkulasi, Work Life Balance Score. Survei ini mengevaluasi bagaimana kita menjalani hidup, baik dalam kehidupan profesional maupun

kehidupan pribadi: survei ini mencerminkan seberapa baik Anda membentuk gaya hidup, kebiasaan, dan perilaku untuk memaksimalkan kepuasan hidup Anda secara keseluruhan dalam lima dimensi berikut:

1. Healthy Body, mencerminkan kebugaran dan kebiasaan sehat Anda;
2. Healthy Mind, menunjukkan seberapa baik Anda menerima emosi positif;
3. Expertise, mengukur kemampuan untuk mengembangkan keahlian dan mencapai sesuatu yang unik;
4. Connection, menilai kekuatan jaringan sosial Anda dan kecenderungan Anda untuk menemukan dunia;
5. Meaning, mengevaluasi kasih sayang, kemurahan hati Anda dan seberapa besar 'Anda menjalani kehidupan impian Anda'

2.2 Pre-processing Data

Pada tahap ini, dilakukan pemilihan atribut yang digunakan, pembersihan data, analisa atribut dan transformasi data. Tahap pertama pemilihan atribut, dari 24 atribut yang ada dihapus atribut timestamp, lalu dilakukan seleksi fitur dengan Principal Component Analysis (PCA) dengan penyesuaian jumlah komponen yang mencakup sekitar 90% varians kumulatif, sehingga atribut yang digunakan berjumlah 20 atribut (menghilangkan atribut timestamp, Age, Gender dan Work Life Balance. Atribut yang dipilih akan disajikan pada gambar 2.

```
df.info()
<class 'pandas.core.frame.DataFrame'>
Int64Index: 15971 entries, 0 to 15971
Data columns (total 20 columns):
#   Column                Non-Null Count  Dtype  
---  --
0   FRUITS_VEGGIES         15971 non-null  int64  
1   DAILY_STRESS           15971 non-null  int32  
2   PLACES_VISITED         15971 non-null  int64  
3   CORE_CIRCLE            15971 non-null  int64  
4   SUPPORTING_OTHERS      15971 non-null  int64  
5   SOCIAL_NETWORK         15971 non-null  int64  
6   ACHIEVEMENT            15971 non-null  int64  
7   DONATION               15971 non-null  int64  
8   BMI_RANGE              15971 non-null  int64  
9   TODO_COMPLETED         15971 non-null  int64  
10  FLOW                   15971 non-null  int64  
11  DAILY_STEPS            15971 non-null  int64  
12  LIVE_VISION            15971 non-null  int64  
13  SLEEP_HOURS            15971 non-null  int64  
14  LOST_VACATION          15971 non-null  int64  
15  DAILY_SHOUTING         15971 non-null  int64  
16  SUFFICIENT_INCOME      15971 non-null  int64  
17  PERSONAL_AWARDS        15971 non-null  int64  
18  TIME_FOR_PASSION        15971 non-null  int64  
19  WEEKLY_MEDITATION      15971 non-null  int64  
dtypes: int32(1), int64(19)
memory usage: 2.5 MB
```

Gambar 2. Fitur yang digunakan

Pada proses pembersihan data, ditemukan 1 data yang kemungkinan input error sehingga penulis memutuskan untuk menghapus data tersebut. Selanjutnya tahap Analisa atribut, atribut awalnya memiliki 1 variabel kategorikal (Daily Stress) yang kemudian diubah menjadi variabel numerik sehingga menjadi 20 variabel numerik, .

2.3 Modeling

Dalam proses ini, dataset dibagi menjadi dua bagian: data training dan data testing. Data training digunakan untuk membuat model klasifikasi, sementara data testing digunakan untuk mengevaluasi model yang telah dibuat. Pembagian dataset ini dilakukan dengan proporsi 80% untuk data training dan 20% untuk data testing.

Pada penelitian ini dibandingkan antara efektifitas algoritma machine learning SVM dan juga random forest. Dengan algoritma svm dan random forest, akan dilakukan proses klasifikasi. Lalu selanjutnya akan ditampilkan visualisasi fitur penting dan menampilkan classification report dari model yang telah dibuat.

2.3.1 Support Vector Machine

SVM adalah algoritma yang bekerja pada hyperplane yang berguna untuk menyortir ilustrasi baru (hyperplane adalah garis yang memisahkan bidang menjadi dua bagian dimana setiap kelas terletak di salah satu sisinya. Karena penggunaannya yang mudah diterapkan (fleksibel) dan mempunyai performa yang cukup baik, sehingga algoritma SVM banyak diterapkan di berbagai bidang (Imaduddin et al., 2021). Untuk kernel yang akan digunakan untuk penelitian adalah rbf kernel dan linear kernel dengan beberapa parameter C.

2.3.2 SVM dengan RBF Kernel

RBF Kernel dapat digunakan saat pengetahuan yang diperoleh dari data yang akan diproses sangat rendah / bahkan tidak ada. Persamaan yang digunakan untuk 2 sampel x dan y adalah sebagai berikut

$$k(x, y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right) \quad (1)$$

2.3.3 SVM dengan Linear Kernel

Kernel Linear digunakan ketika data dapat dipisahkan secara linear, yaitu dapat dipisahkan dengan satu garis lurus. Ini adalah jenis kernel yang paling umum digunakan, terutama ketika kumpulan data memiliki banyak fitur. Persamaan yang digunakan untuk dua sampel, x dan y, adalah sebagai berikut:

$$k(x, y) = x^T \cdot y \quad (1)$$

2.3.4 Random Forest

Random Forest adalah algoritma yang mempertimbangkan banyak pohon keputusan, sehingga membentuk sebuah hutan disebut juga ansambel algoritma pohon keputusan (Ahuja & Banga, 2019). Random Forest banyak digunakan karena kelebihan nya pada akurasi yang lebih baik, dan ketahanan dari berbagai gangguan, kecepatan dan kemudahan dalam implementasi model (Goel, 2017; Jollyta et al., 2019)

2.4 Evaluasi Model

Setelah model telah selesai dibuat, maka akan dilakukan evaluasi terhadap model yang telah dibuat dengan melakukan perbandingan dari kedua jenis algoritma tersebut untuk menentukan mana algoritma yang cocok digunakan untuk model yang telah dibuat, Setelah menentukan model mana yang akan digunakan, maka selanjutnya membuat model yang dapat menangani input user

2.5 Deploy Model

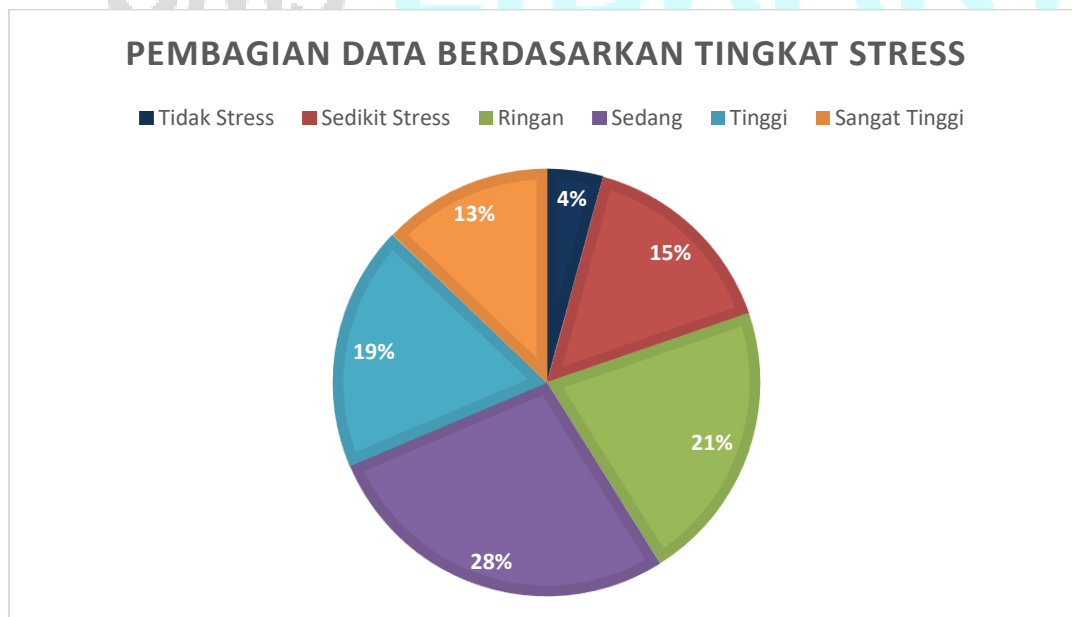
Pada tahap ini, akan dilakukan deployment dengan menggunakan streamlit. Streamlit adalah sebuah plugin yang digunakan untuk membuat visualisasi situs web dan analisis data untuk tujuan dashboard menggunakan bahasa pemrograman Python(Streamlit, 2023)

2.6 Pengujian

Dalam tahapan pengujian, metode System Usability Scale (SUS) akan digunakan untuk mengidentifikasi kekurangan atau kesalahan yang mungkin ada pada sistem yang dibuat. (Renaldi & Aryo, 2020). SUS merupakan metode yang digunakan untuk menilai produk berdasarkan kegunaan produk atau layanan tersebut.

3. HASIL DAN PEMBAHASAN

Dari data yang berasal dari Kaggle berjumlah sebanyak 15.971, pembagian data terhadap Target (Tingkat stress), sebanyak 676 data berlabel 0 (Tidak Stress), 2478 data berlabel 1 (Sedikit Stress), 3407 berlabel 2 (Ringan), 4398 berlabel 3 (Sedang), 2960 berlabel 4(Tinggi), 2052 berlabel 5(Sangat Tinggi). Lebih jelas akan ditampilkan pada Gambar 3 dibawah ini.



Gambar 3. Pembagian data berdasarkan Tingkat Stress

3.1. Pembuatan Model

Berdasarkan data yang diperoleh, dilakukan beberapa percobaan untuk merancang model untuk penelitian ini dengan menggunakan Support Vector Machine dan Random Forest

3.1.1. Percobaan ke-1

Dilakukan pengujian menggunakan model SVM dan RF, dengan menggunakan data yang sudah dimiliki tanpa ada perubahan data apapun. Dari pengujian pada percobaan ini, diperoleh hasil yang disajikan pada tabel 2.

Tabel 2. Percobaan ke-1 (Tanpa perubahan data)

Model	SVM default	RBF dan C= 100	RBF dan C=1000	Linear dan C=1.0	Random Forest
Skor Akurasi	0.3280	0.2926	0.2920	0.3030	0.3527

Dari hasil diatas, skor akurasi ada di angka 0.30 – 0.35. Angka tersebut sangat rendah jika ingin digunakan sebagai model machine learning. Perlu peningkatan efektivitas model karena Tingkat akurasi yang didapat sangat rendah.

3.1.2. Percobaan ke-2

Pada percobaan ke-2, akan dilakukan pengelompokan kategori pada target menjadi 2 kategori, yaitu “Stress” dan “Tidak Stress”. Untuk pengelompokan, akan dikelompokkan “0 – 2” masuk ke kategori “Tidak Stress”, dan “3 – 5” masuk ke kategori “Stress”. Sehingga jumlah data menjadi 6.561 data berlabel “Tidak Stress” dan 9.410 data berlabel “Stress”. Setelah melakukan pengelompokan kategori maka didapatkan hasil pengujian yang dapat dilihat pada tabel 3 dibawah ini

Tabel 3. Percobaan ke-2 (Pengelompokan Kategori target)

	SVM default	RBF dan C= 100	RBF dan C=1000	Linear dan C=1.0	Random Forest
Skor Akurasi	0.6632	0.6166	0.6085	0.6613	0.6732

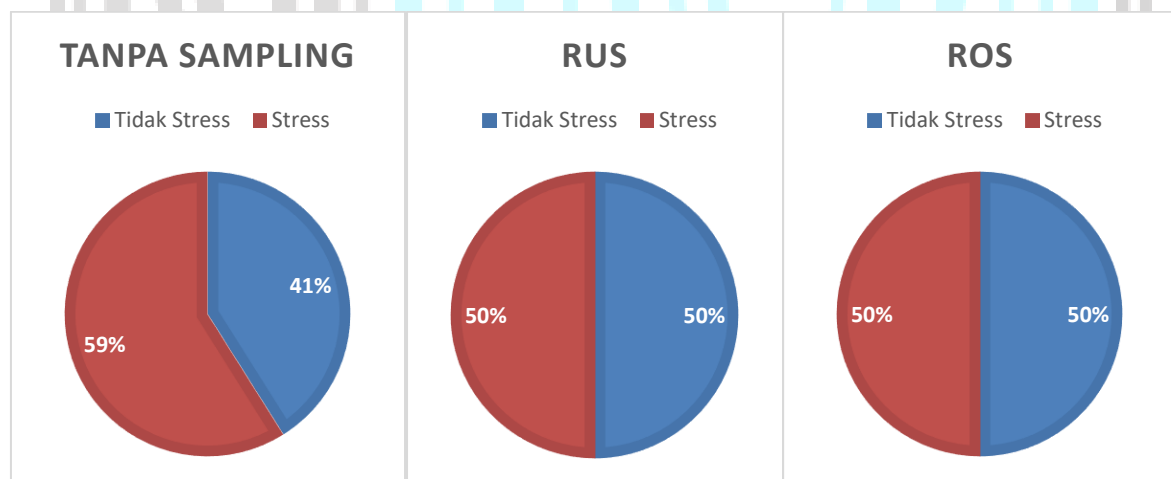
Pada tabel 2, dapat dilihat jika model yang ada memiliki Tingkat akurasi yang sudah meningkat dari percobaan pertama, namun menurut beberapa pakar machine learning model machine learning yang bagus biasanya memiliki Tingkat akurasi 70 % keberhasilan.

3.1.3. Percobaan ke-3

Pada Percobaan ke-2, dari pengelompokan Tingkat stress memiliki jumlah data yang tidak seimbang, oleh sebab itu penulis melakukan percobaan ke-3 dengan pendekatan teknik

sampling Random Under Sampling (RUS) dan Random Over Sampling (ROS) untuk mengatasi masalah data tidak berimbang. RUS adalah teknik yang mengurangi data mayoritas dengan cara menghapus beberapa data dari kelas mayoritas, sedangkan ROS adalah teknik sampling yang memperbanyak data minoritas dengan cara menggandakan data data yang tergolong minoritas.(Gumelar et al., 2021; Johnson & Khoshgoftaar, 2019)

Dengan menggunakan RUS, jumlah data mayoritas dikurangi agar sama dengan jumlah data minoritas yaitu 6.561 data tiap kategori, sehingga jumlah data menjadi 13.122 data. Sedangkan saat menggunakan ROS, jumlah data minoritas ditambah (digandakan) agar sama dengan jumlah data mayoritas yaitu 9.410 data tiap kategori, sehingga jumlah data menjadi 18.820 data. Perubahan jumlah data dapat dilihat pada gambar 4. Setelah dilakukan Sampling data menggunakan RUS dan ROS selanjutnya dilakukan pengujian model yang hasilnya dapat dilihat pada Tabel 4 dan Tabel 4.



Gambar 4. Perbandingan Jumlah data Sebelum dan Sesudah dilakukan Teknik Sampling

Tabel 4. Tingkat Akurasi setelah dilakukan Random UnderSampling

	SVM default	RBF dan C= 100	RBF dan C=1000	Linear dan C=1.0	Random Forest
Skor Akurasi	0.6521	0.6228	0.5881	0.6483	0.6716

Tabel 5. Tingkat Akurasi setelah dilakukan Random OverSampling

	SVM default	RBF dan C= 100	RBF dan C=1000	Linear dan C=1.0	Random Forest
Skor Akurasi	0.6700	0.6859	0.6947	0.6522	0.7648

Dari hasil Percobaan yang menggunakan model Random Forest dan menggunakan pendekatan teknik sampling ROS memiliki hasil akurasi yang tertinggi yaitu 0.7648. Untuk hasil selengkapnya disajikan pada tabel 6 dibawah.

Tabel 6. Perbandingan Tingkat akurasi

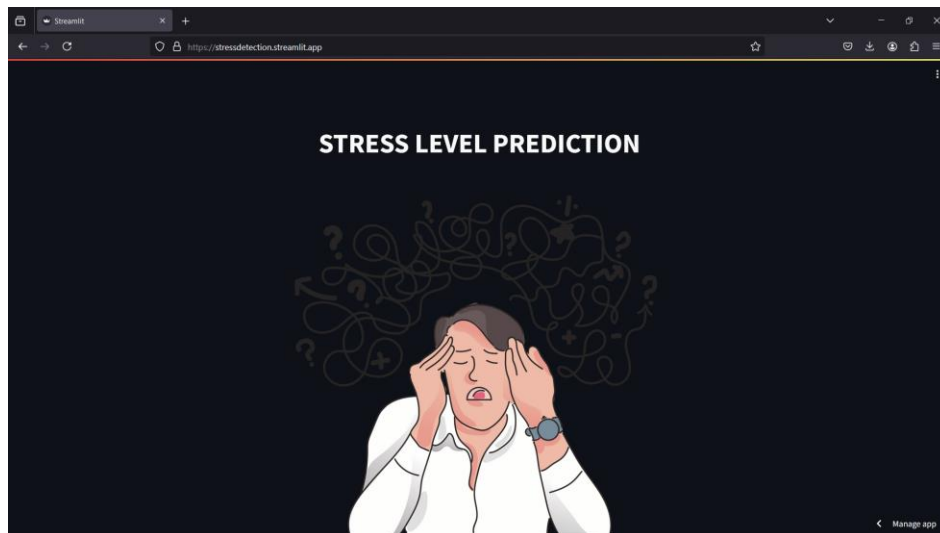
	SVM default	RBF dan C= 100	RBF dan C=1000	Linear dan C=1.0	Random Forest
Default	0.3280	0.2926	0.2920	0.3030	0.3527
Pengelompokan Kategori	0.6657	0.6169	0.6125	0.6660	0.6732
Pengelompokan Kategori + RUS	0.6521	0.6228	0.5881	0.6483	0.6716
Pengelompokan Kategori + ROS	0.6700	0.6859	0.6947	0.6522	0.7648

Berdasarkan data hasil pengujian pada tabel 5, akurasi tinggi sebesar 0,7648 yang dicapai oleh model Random Forest dikombinasikan dengan pengurangan kategori dan Random Over Sampling (ROS) dapat dikaitkan dengan penanganan ketidakseimbangan kelas yang efektif dan ketahanan algoritma Random Forest. Dengan mengurangi jumlah kategori, model dapat fokus pada perbedaan paling penting antara kedua kategori, sehingga model dapat menangkap pola dasar data dengan lebih baik. Selain itu, teknik ROS membantu menyeimbangkan kumpulan data dengan meningkatkan representasi kelas minoritas, yang sangat penting untuk meningkatkan performa model pada kelas minoritas. Model Random Forest, yang dikenal karena kemampuannya menangani interaksi kompleks dan noise dalam data, sangat cocok untuk memanfaatkan kumpulan data seimbang ini, sehingga menghasilkan akurasi tinggi sebesar 0,7667. Berdasarkan tingkat akurasi yang diperoleh, penulis memilih untuk menggunakan model RF + pendekatan ROS sebagai model yang akan dipakai untuk membuat situs web menggunakan streamlit.

3.2. Hasil Deploy streamlit

3.2.1. Halaman Awal

Pada halaman awal digunakan sebagai landing page, pada halaman ini terdapat judul aplikasi dan juga tombol untuk memulai tes untuk memprediksi tingkat stress user. Tampilan dapat dilihat pada gambar 5.

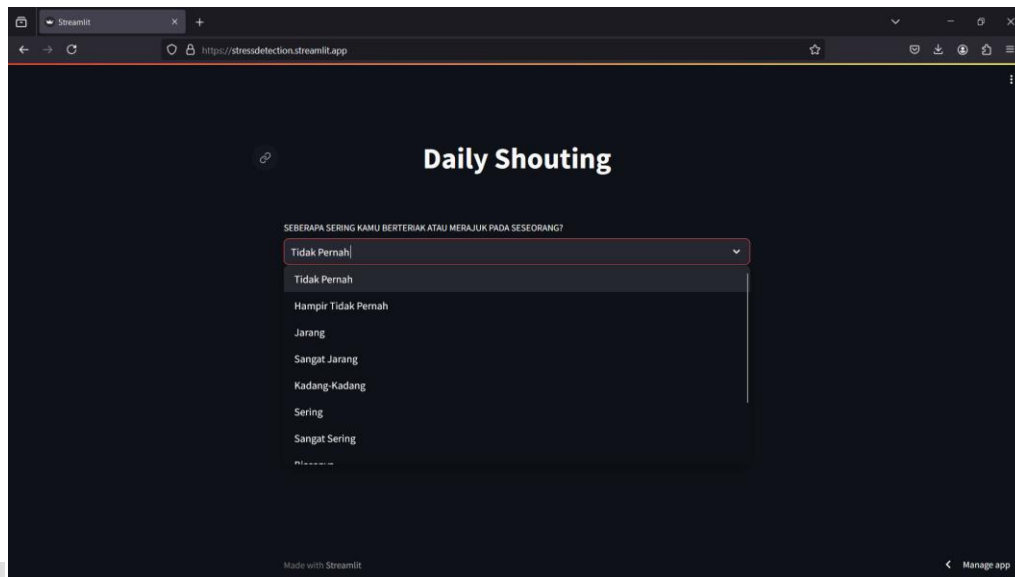


Gambar 5. Halaman Awal

3.2.2. Halaman Pengisian Data

Setelah menekan tombol "Mulai" pada halaman awal, pengguna akan diarahkan ke halaman pengisian data yang berisi 19 pertanyaan. Halaman ini juga dilengkapi dengan tombol "Kembali" yang berguna untuk kembali ke halaman sebelumnya jika terdapat kesalahan dalam pengisian data, serta tombol "Selanjutnya" untuk melanjutkan ke pertanyaan berikutnya. Pada akhir proses pengisian data, tombol "Selanjutnya" akan berubah menjadi tombol "Prediksi". Tampilan halaman pengisian data ini dapat dilihat pada Gambar 6 dan Gambar 7.

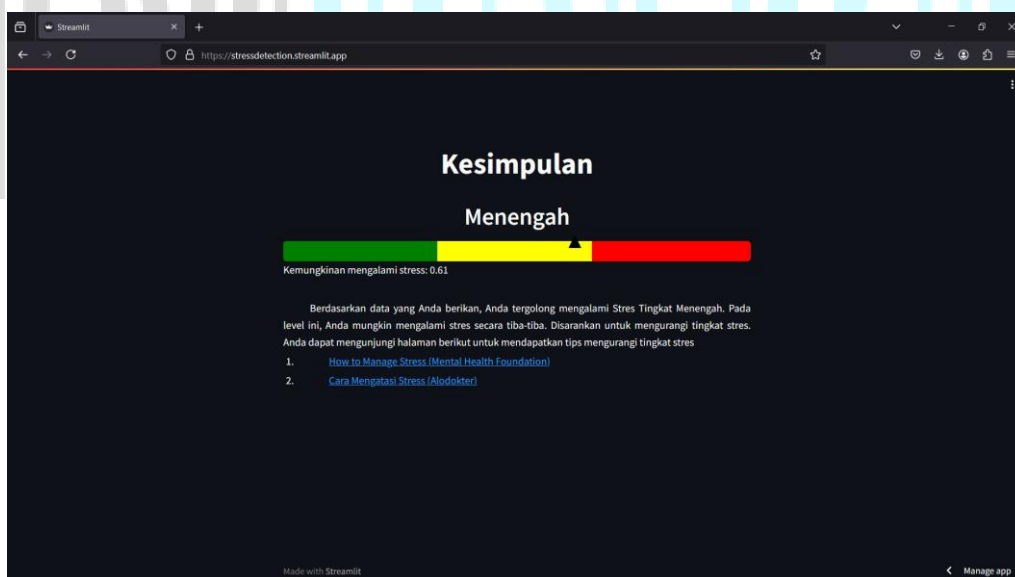
Gambar 6. Halaman Pengisian data



Gambar 7. Halaman Pengisian Data Terakhir

3.2.3. Halaman penampilan hasil prediksi

Gambar 8 menampilkan hasil prediksi setelah user selesai menjawab pertanyaan yang ada. Pada halaman ini menampilkan range tingkat stress berdasarkan probabilitas yang telah diolah dengan machine learning, dan juga beberapa situs yang memberikan tips tips untuk mengurangi tingkat stress seseorang.



Gambar 8. Halaman Hasil Prediksi

3.3. Pengujian System Usability Scale

Pengujian dengan metode SUS dilakukan pembagian kuesioner gform yang memiliki 10 pertanyaan terkait tanggapan responden ketika mengakses website. Sebelum responden mengisi kuesioner, responden mengakses website untuk mencoba fitur – fitur yang ada pada website. Pengujian melibatkan 30 responden dari kalangan umum. Hasil pengujian dapat dilihat pada tabel 7.

Tabel 7. Hasil penilaian kuesioner

No	Responden	PERTANYAAN										TOTAL	Total x 2,5
		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10		
1	R 1	4	1	5	1	5	2	5	2	4	2	31	77,5
2	R 2	3	4	3	3	4	2	3	2	3	4	31	77,5
3	R 3	5	3	4	2	4	2	4	2	4	3	33	82,5
4	R 4	5	1	5	1	5	1	5	1	5	1	30	75
5	R 5	3	2	4	1	4	2	4	2	4	2	28	70
6	R 6	4	4	4	4	4	2	4	4	4	1	35	87,5
7	R 7	4	3	2	2	4	2	4	3	5	1	30	75
8	R 8	5	1	5	1	5	2	5	1	5	1	31	77,5
9	R 9	3	4	2	5	3	3	4	4	3	5	36	90
10	R 10	3	3	3	2	3	3	3	2	4	2	28	70
11	R 11	3	2	3	3	3	3	3	3	3	3	29	72,5
12	R 12	4	1	5	1	4	3	4	1	5	1	29	72,5
13	R 13	5	1	5	1	4	1	5	1	5	1	29	72,5
14	R 14	3	4	3	2	3	2	4	3	3	5	32	80
15	R 15	3	2	4	1	4	2	4	1	4	1	26	65
16	R 16	3	1	5	1	5	1	5	1	5	1	28	70
17	R 17	5	1	5	1	5	1	5	1	5	1	30	75
18	R 18	4	1	5	2	4	1	5	1	4	2	29	72,5
19	R 19	1	2	3	4	5	1	2	3	4	5	30	75
20	R 20	5	1	5	1	5	1	5	1	5	3	32	80
21	R 21	4	3	3	2	4	2	5	1	4	4	32	80
22	R 22	3	1	5	1	5	1	5	2	5	1	29	72,5
23	R 23	4	3	3	1	4	2	4	2	4	2	29	72,5
24	R 24	4	2	4	1	4	2	4	1	4	1	27	67,5
25	R 25	4	1	5	1	4	2	4	4	5	2	32	80
26	R 26	3	3	4	2	4	2	4	2	3	5	32	80
27	R 27	5	1	5	1	4	2	5	1	5	2	31	77,5
28	R 28	3	2	4	2	3	3	4	2	4	3	30	75
29	R 29	4	2	4	2	4	2	4	2	4	4	32	80
30	R 30	3	3	4	5	3	3	4	4	4	4	37	92,5
Jumlah												918	2295
Rata Rata Hasil Akhir													76,5

Deskripsi pertanyaan:

P1: Saya berencana untuk terus menggunakan sistem ini di masa mendatang.

P2: Saya mengalami kesulitan dalam menggunakan sistem ini.

P3: Saya merasa sistem ini mudah digunakan.

P4: Saya memerlukan bantuan orang lain saat menggunakan sistem ini.

P5: Saya merasa berbagai menu dalam sistem ini berfungsi dengan baik.

P6: Saya merasa ada banyak hal yang tidak konsisten dalam sistem ini.

P7: Saya yakin orang lain akan cepat memahami cara menggunakan sistem ini.

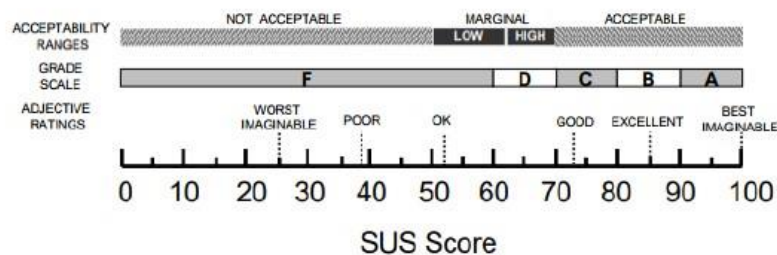
P8: Saya merasa sulit memahami cara kerja sistem ini.

P9: Saya dapat menggunakan sistem ini dengan lancar.

P10: Saya perlu beradaptasi sebelum bisa menggunakan sistem ini.

Keterangan jawaban:

- 1 : Sangat Tidak Setuju
- 2 : Tidak Setuju
- 3 : Ragu – ragu
- 4 : Setuju
- 5 : Sangat Setuju



Gambar 9. Rentang Kategori Pengujian

Berdasarkan hasil pengujian dengan metode SUS pada tabel 6 didapatkan skor rata-rata 76,5. Berdasarkan hasil pengujian dan Gambar 9 sebagai acuan penilaian dapat disimpulkan bahwa system yang dikembangkan termasuk dalam kategori baik atau Good and Acceptable dengan grade scale C.

4. PENUTUP

Berdasarkan apa yang telah dilakukan pada penelitian melalui proses yang telah dijelaskan pada metodologi serta hasil dan pembahasan, penulis mendapatkan beberapa Kesimpulan diantaranya bahwa aplikasi “Stress Detection” telah bekerja sesuai dengan yang diharapkan dan telah masuk kedalam kategori baik atau Good and Acceptable dengan grade scale C. Pengembangan dari aplikasi masih memungkinkan agar system dapat bekerja lebih baik dan memiliki fitur yang lebih lengkap. Beberapa saran pengembangan aplikasi, diantaranya percobaan model bisa dengan mengganti algoritma yang lebih baik atau pengembangan dari model yang sudah ada, dengan harapan tingkat akurasi model yang diperoleh lebih maksimal. Diharapkan aplikasi ini dapat bermanfaat bagi semua pengguna dalam membantu untuk mengetahui tingkat stress pengguna lebih awal dan memberikan tips untuk mengurangi tingkat stress.

DAFTAR PUSTAKA

- Ahuja, R., & Banga, A. (2019). Mental stress detection in university students using machine learning algorithms. *Procedia Computer Science*, 152, 349–353. <https://doi.org/10.1016/j.procs.2019.05.007>
- Dalat, Y. (2021). *Lifestyle and Wellbeing Data*. <https://www.kaggle.com/datasets/ydalat/lifestyle-and-wellbeing-data>
- Fadhila, A. Van, Azzahra, J. A., Rizki, K., Zulkarnain, T., Lathifah, N. D., Salsabiila, S. Z., & Chamidah, N. (2023). Implementasi Metode Machine Learning Untuk Mendeteksi Tingkat Stres Manusia Berdasarkan Kualitas Tidur. *Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer Dan Aplikasinya*, 4(1), 130–143.
- Goel, E. et al. (2017). Random forest: A review. *International Journal of Advanced Research in Computer Science and Software Engineering*, 7(1), 251–257.
- Gumelar, G., Ain, Q., Marsuciati, R., Agustanti Bambang, S., Sunyoto, A., & Syukri Mustafa, M. (2021). Kombinasi Algoritma Sampling dengan Algoritma Klasifikasi untuk Meningkatkan Performa Klasifikasi Dataset Imbalance. *Prosiding SISFOTEK*, 5(1), 250–255. <http://www.seminar.iaii.or.id/index.php/SISFOTEK/article/view/295>
- Hanchate, R., Narute, H., & Shavage, S. (2023). Stress Detection Using Machine Learning. *International Journal of Science and Healthcare Research*, 8(2), 307–311. <https://doi.org/10.52403/ijshr.20230239>
- Imaduddin, H., Aditya Hermansyah, B., & Aura Salsabilla, F. B. (2021). COMPARISON OF SUPPORT VECTOR MACHINE AND DECISION TREE METHODS IN THE CLASSIFICATION OF BREAST CANCER. *Jurnal Pendidikan Teknologi Informasi*, 5(1), 22–30.
- Johnson, J. M., & Khoshgoftaar, T. M. (2019). Deep Learning and Data Sampling with Imbalanced Big Data. *IEEE*, 175–183. <https://doi.org/10.1109/IRI.2019.00038>
- Jollyta, D., Sukrianto, D., & Mahaputra Riau Pekanbaru, A. (2019). Analysis of Slow Moving Goods Classification Technique: Random Forest and Naïve Bayes. *Khazanah Informatika: Jurnal Ilmu Komputer Dan Informatika*, 5(2), 134–139.
- Kemenkes. (2013). *Laporan Riskesdas 2013 Nasional*.
- Kemenkes. (2018). *Laporan Riskesdas 2018 Nasional*.
- Lalenoh, G. A., Berkat N Zega, I. P., Febe Yuni, I., Ayu Florensa, M. V, Tri Anggraini N, M. S., Fakultas Keperawatan, M., Pelita Harapan, U., & Fakultas Keperawatan, D. (2021). HUBUNGAN TINGKAT STRES DENGAN IDE BUNUH DIRI PADA MAHASISWA THE RELATIONSHIP BETWEEN STRESS LEVELS AND SUICIDE IDEATION IN COLLEGE STUDENTS. *Nursing Current*, 9(1), 89–101.
- Maruyama, S., & Morimoto, K. (1997). The Effects of Lifestyle and Type A Behavior on the Life-Stress Process. *Environmental Health and Preventive Medicine*, 2, 28–34.
- National Adolescent Mental Health Survey. (2022). *I-NAMHS: Indonesia-National Adolescent Mental Health Survey*.
- Omasu, F., Kawano, A., Nagayasu, M., & Nishi, A. (2022). Research on Lifestyle Habits Caused by Stress. *Open Journal of Preventive Medicine*, 12(09), 190–198. <https://doi.org/10.4236/ojpm.2022.129014>
- Renaldi, R., & Aryo, D. (2020). Sistem Informasi Geografis Pemetaan Sekolah Menengah Atas / Sederajat di Kota Surakarta Menggunakan Leaflet Javascript Library Berbasis Website. *Jurnal Teknik Elektro*, Vol .20(No. 2), 112.
- Streamlit. (2023). *Streamlit Docs*. <https://docs.streamlit.io>
- Ulfa, L., & Rizqi, M. (2019). Faktor Penyebab Stress dan Dampak Bagi Kesehatan. *OSF Preprints*. <https://doi.org/https://doi.org/10.31219/osf.io/h4cnv>
- Zamzani, Z. M., Ryan Nurdiansyah, M. N., Yudha Yana, B., & Artikel, S. (2023). DETEKSI STRES MANUSIA MELALUI ANALISIS TIDUR DENGAN METODE FUZZY

INFO ARTIKEL ABSTRACT. *Industri Dan Informasi*, 1(1), 58–71.
<https://journal.awatarapublisher.com/index.php/technovatar>

