

PREDIKSI TINGKAT KESEPIAN BERDASARKAN PROFIL MEDIA SOSIAL MENGUNAKAN ALGORITMA *RANDOM FOREST*

Asep Lukman Arip Hidayat¹, Achmad Solichin²

^{1,2} Magister Ilmu Komputer Fakultas Teknologi Informasi Universitas Budi Luhur Jakarta
Jl. Ciledug Raya, Petukangan Utara, Kec. Pesanggrahan, Kota Jakarta Selatan, DKI Jakarta

¹ aseplukman.ah@gmail.com

² achmad.solichin@budiluhur.ac.id

Abstract

Pada era digital yang semakin berkembang, media sosial telah menjadi bagian penting yang tidak terpisahkan dari kehidupan manusia modern, terutama bagi generasi muda, termasuk Generasi Z di SMK Tunas Media yang lahir pada awal abad ke-21 dan menjadi generasi pertama yang tumbuh dengan akses internet, teknologi dan media sosial. Beberapa penelitian terkini mengungkapkan bahwa meskipun Generasi Z tumbuh dalam era teknologi yang sangat terhubung, mereka ternyata memiliki gejala tingkat kesepian yang sangat tinggi bahkan melampaui seluruh generasi sebelumnya, salah satu penyebabnya adalah penggunaan media sosial yang berlebihan, sehingga mereka lebih mengutamakan komunikasi digital dibandingkan interaksi secara langsung. Sekitar 73% dari mereka mengatakan bahwa mereka merasa terisolasi. Hal ini tentunya berdampak pada kesehatan mental mereka, 91% melaporkan mengalami stres fisik atau emosional, dan 68% mengatakan bahwa mereka merasa khawatir tentang masa depan. Oleh karena itu, sangatlah penting untuk melakukan deteksi dini pada siswa-siswi SMK Tunas Media yang berisiko mengalami kesepian. Teknologi pembelajaran mesin *Random Forest*, dapat digunakan untuk memprediksi tingkat kesepian siswa berdasarkan data profil media sosial. Algoritma ini dapat mempelajari data profil media sosial, seperti pola aktivitas, konteks konten yang dibagikan, interaksi sosial, ekspresi emosi, dan jaringan sosial. Selain itu, penggunaan data *UCLA Loneliness Scale* merupakan salah satu alat psikometri yang paling umum digunakan untuk mengukur tingkat kesepian dapat diintegrasikan dengan hasil prediksi model pembelajaran mesin untuk memberikan validitas tambahan dalam menilai kesepian. Penelitian ini menunjukkan hasil yang signifikan dalam peningkatan kinerja model klasifikasi setelah diterapkan metode *Random Forest* yang dioptimalkan dengan *Optuna*. Dalam penelitian ini, model digunakan untuk memprediksi tingkat kesepian berdasarkan profil media sosial, dan berhasil mencapai akurasi 90%. Hal ini menunjukkan bahwa model *Random Forest* yang dioptimalkan dengan *Optuna* memiliki potensi besar dalam mengklasifikasikan tingkat kesepian rendah dan tinggi pada siswa generasi Z.

Keywords: Kesepian, Loneliness, Random Forest, Social Media, Z generation.

I. PENDAHULUAN

SMK Tunas Media adalah sekolah menengah kejuruan yang berlokasi di Kota Depok, Jawa Barat dengan konsentrasi keahlian Rekayasa Perangkat Lunak, Desain Komunikasi Visual, dan Manajemen Perkantoran. Penerapan Sistem Informasi Manajemen sekolah di lingkungan SMK Tunas Media telah mendorong penggunaan perangkat digital seperti ponsel dan laptop dalam proses pembelajaran. Hal ini menjadi sangat relevan bagi siswa-siswi SMK Tunas Media yang tidak hanya menggunakan teknologi untuk meningkatkan produktivitas pembelajarannya tetapi juga dapat berinteraksi di dunia maya atau media sosial, sehingga tidak menutup kemungkinan mereka menghabiskan waktu cukup banyak di internet.

Pada era digital yang semakin berkembang, media sosial telah menjadi bagian penting yang tidak terpisahkan dari

kehidupan manusia modern, terutama bagi generasi muda termasuk Generasi Z, yaitu kelompok orang-orang yang lahir dari tahun 1997 hingga 2012, termasuk siswa-siswi SMK Tunas Media adalah bagian dari generasi ini yang lahir pada awal abad ke-21 dan menjadi generasi pertama yang tumbuh dengan akses internet, teknologi dan media sosial, sehingga mereka lebih banyak berinteraksi melalui *platform online* dibanding interaksi secara langsung. Generasi Z dan *milenial* memiliki *preferensi* media sosial yang berbeda sesuai data yang dikumpulkan dari survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) sebanyak 8.720 orang yang dipilih secara acak dan proporsional dari 38 provinsi di Indonesia [1]. Generasi Z Indonesia mayoritas (51,9%) sering menggunakan aplikasi media sosial Instagram. Di sisi lain, mayoritas *milenial* (74,09%) sering menggunakan Facebook. Tidak hanya itu, penggunaan media sosial juga mempengaruhi cara mereka berkomunikasi, berinteraksi dan mengonsumsi informasi,

sehingga media sosial bisa berpotensi memiliki efek negatif yang dapat mempengaruhi kesehatan mental siswa di sekolah, terutama tingkat kesepian.

Menurut [2] kesepian adalah perasaan tidak menyenangkan yang muncul ketika seseorang merasa kekurangan hubungan sosial yang bermakna. Pengalaman ini bersifat subjektif, sehingga seseorang dapat merasa kesepian meskipun berada di keramaian. Sebaliknya ada pula orang yang meskipun sendirian, tidak mengalami rasa kesepian. Kesepian sering dikaitkan dengan berbagai konsekuensi negatif, termasuk masalah kesehatan mental seperti depresi atau stres, serta masalah kesehatan fisik seperti penyakit kronis.

Studi terkini menunjukkan bahwa meskipun Generasi Z tumbuh dalam era keterhubungan digital, mereka justru mengalami tingkat kesepian lebih tinggi dibanding generasi sebelumnya. Penggunaan media sosial yang berlebihan mendorong pergeseran dari interaksi tatap muka menuju komunikasi digital, di mana *like*, komentar, dan *share* dijadikan tolok ukur hubungan sosial. Kondisi ini berkontribusi pada meningkatnya perasaan terisolasi; sekitar 73% Generasi Z mengaku kesepian, 91% mengalami stres fisik maupun emosional, dan 68% merasa cemas tentang masa depan [3]. Media sosial sendiri berfungsi sebagai sarana untuk berbagi informasi, foto, dan ide tanpa batas geografis. Platform seperti Facebook, Instagram, Twitter, dan TikTok memungkinkan terbentuknya komunitas daring yang menjaga keterhubungan sosial meski pengguna berada di lokasi berbeda. Dengan demikian, media sosial memainkan peran penting dalam interaksi sosial modern, sekaligus berpotensi memengaruhi kesejahteraan psikologis penggunanya [4].

Berdasarkan survei yang dilakukan oleh *Into The Light Indonesia Bersama Change.org Indonesia* pada periode 25 Mei -16 Juni 2021, sebanyak 5.106 orang (98%) dari 5.211 responden melaporkan bahwa mereka mengalami kesepian dalam empat minggu terakhir. Tingkat kesepian tertinggi ditemukan pada kelompok usia 18-24 tahun, yang sebagian besar termasuk Generasi Z, dengan nilai prevalensinya 0,987, disusul kelompok usia 25-34 tahun sebesar 0,985, dan kelompok usia ≥ 35 tahun dengan 0,951. Mayoritas responden adalah berusia 18-24 tahun (45,9%), disusul oleh kelompok usia 25-34 tahun (34,0%), dan sisanya merupakan mereka yang berusia 35 tahun ke atas (20,1%). Sebagian besar responden tinggal di Pulau Jawa (75,8%), sementara 24,2% lainnya berasal dari luar Jawa. Kemudian dari segi pendidikan, lebih dari separuh peserta (54,4%) memiliki gelar diploma atau sarjana, sedangkan 45,6% hanya menyelesaikan pendidikan SMA atau lebih rendah [5].

Penelitian sebelumnya menunjukkan adanya keterbatasan dalam mendeteksi tingkat kesepian secara dini melalui data profil media sosial. Sebagian besar pendekatan yang digunakan masih mengandalkan metode kualitatif, seperti survei dan wawancara, yang cenderung subjektif serta membutuhkan waktu lebih lama. Hal ini menegaskan perlunya metode yang lebih objektif, efisien, dan adaptif terhadap perkembangan teknologi. Salah satu pendekatan yang relevan adalah

penerapan pembelajaran mesin, khususnya algoritma *Random Forest*, yang mampu mengidentifikasi pola perilaku media sosial mencakup aktivitas pengguna, jenis konten yang dibagikan, interaksi sosial, ekspresi emosional, hingga struktur jaringan sosial.

Dalam penelitian ini, algoritma *Random Forest* diusulkan sebagai metode prediksi tingkat kesepian pada siswa-siswi SMK Tunas Media dengan memanfaatkan data profil media sosial. Model ini bekerja dengan membangun sejumlah pohon keputusan, kemudian menghasilkan prediksi akhir berdasarkan agregasi dari seluruh pohon tersebut. Keunggulan utama *Random Forest* terletak pada kemampuannya mengolah data dalam skala besar dan kompleks sekaligus meminimalkan risiko *overfitting*, yang sering menjadi kelemahan pada penggunaan pohon keputusan tunggal [6]. Sejumlah penelitian menunjukkan bahwa pembelajaran mesin dapat dimanfaatkan untuk menganalisis data media sosial dalam mengidentifikasi kondisi psikologis, termasuk kesepian [7]. Algoritma seperti *Support Vector Machine (SVM)* dan *Random Forest* mampu mendeteksi kesepian secara otomatis dengan akurasi tinggi melalui analisis pola komunikasi, frekuensi interaksi, dan penggunaan kata pada media sosial [7]. Validitas prediksi dapat diperkuat dengan mengintegrasikan model pembelajaran mesin dengan *UCLA Loneliness Scale*, yang telah terbukti *reliabel* dan *valid* untuk berbagai kelompok usia [8]. Lebih lanjut, penggunaan analisis sentimen dan pengolahan teks dari media sosial yang digabungkan dengan instrumen psikometrik seperti *UCLA Loneliness Scale* dapat meningkatkan akurasi prediksi kesepian [9].

UCLA Loneliness Scale merupakan instrumen psikologis yang dikembangkan oleh [2] dan direvisi menjadi versi ketiga pada tahun 1996. Skala ini banyak digunakan dalam penelitian psikologi sosial dan kesehatan masyarakat karena dinilai praktis serta memiliki reliabilitas dan validitas yang tinggi dalam mengukur dampak kesepian terhadap kesehatan mental maupun fisik [2]. Instrumen ini terdiri atas 20 pernyataan terkait pengalaman sosial responden, yang dijawab menggunakan skala *Likert* 4 poin (1 = Tidak Pernah, 2 = Jarang, 3 = Kadang-kadang, 4 = Sering). Beberapa item bersifat positif dan memerlukan pembalikan skor (*reverse scoring*) agar hasil lebih konsisten. Dalam penelitian terbaru, *UCLA Loneliness Scale* digunakan untuk mengidentifikasi tingkat kesepian dengan rentang skor total 20–80. Batas skor >43 menunjukkan kesepian tinggi, sedangkan skor ≤ 43 menunjukkan kesepian rendah [10]. Proses perhitungan dilakukan dengan menjumlahkan semua item, termasuk skor terbalik untuk pernyataan tertentu. Dengan demikian, skor akhir memberikan gambaran kuantitatif mengenai kondisi kesepian responden serta dapat menjadi dasar intervensi atau dukungan sosial yang diperlukan.

$$\text{Skor Terbalik} = (\text{Jumlah Skala Poin} + 1) - \text{Jawaban Responden} \quad (1)$$

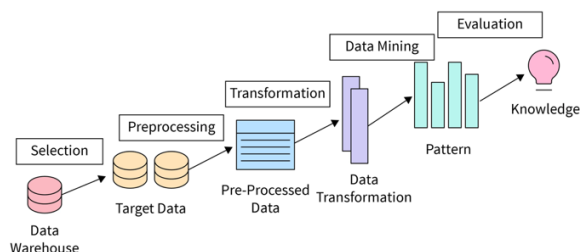
Penelitian ini bertujuan untuk menyediakan model deteksi dini untuk memprediksi tingkat kesepian siswa-siswi SMK Tunas Media dengan menggabungkan data profil media sosial mereka dengan *UCLA Loneliness Scale*. *Random Forest* dipilih karena kemampuan untuk menangani data yang besar dan kompleks serta dapat memberikan fitur penting, yang berguna untuk mengevaluasi hubungan antara data profil media sosial dengan tingkat kesepian Generasi Z di SMK Tunas Media. Penelitian ini dapat menjadi dasar untuk pengembangan alat deteksi dini berbasis pada profil media sosial untuk menemukan siswa-siswi yang berisiko mengalami kesepian sehingga intervensi yang tepat dapat dilakukan lebih awal. Selain itu, penelitian ini memiliki potensi untuk mengungkap lebih banyak informasi tentang hubungan antara perilaku di media sosial dengan kesejahteraan mental.

II. METODOLOGI PENELITIAN

Bagian ini menjelaskan secara rinci tentang penelitian yang dilakukan mulai dari langkah-langkah sampai metode yang digunakan.

A. Metode Penelitian

Penelitian ini menggunakan metode *Knowledge Discovery in Database (KDD)* sebagai pendekatan utama untuk mengolah data dan menemukan pola yang relevan untuk mendukung prediksi tingkat kesepian siswa SMK Tunas Media berdasarkan data profil media sosial mereka.



Gambar 1. *Knowledge Discovery in Databases (KDD)* [11]

Knowledge Discovery in Database (KDD) merupakan proses sistematis yang mencakup beberapa tahapan, mulai dari pengumpulan hingga interpretasi data. Implementasi algoritma *Random Forest* dalam penelitian ini dilakukan melalui beberapa tahapan sebagai berikut:

- 1) *Persiapan Data*: Data hasil kuesioner online yang telah melewati prapemrosesan data digunakan sebagai masukan untuk algoritma *Random Forest*, kemudian variabel-variabel penting, seperti pola aktivitas, jenis konten, dan interaksi sosial, dipilih sebagai fitur utama dalam model.
- 2) *Pembagian Dataset*: Dataset dipecah menjadi dua bagian, yaitu data pelatihan (80%) untuk melatih model dan data pengujian (20%) untuk mengevaluasi performa model.
- 3) *Proses Pelatihan Model*: Model dilatih dengan membangun sejumlah pohon keputusan (*decision trees*) secara acak. Setiap pohon belajar dari *subset data* untuk menemukan pola-pola yang dapat membantu prediksi.

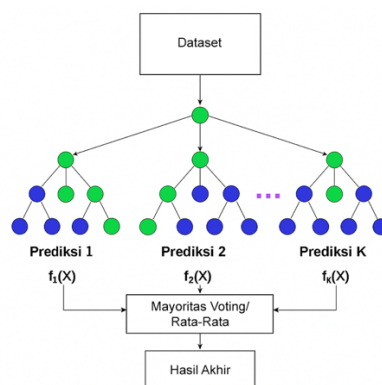
4) *Pengujian Model*: Data pengujian digunakan untuk mengukur kinerja model yang telah dilatih, dengan membandingkan hasil prediksi dengan data aktual.

5) *Evaluasi Kinerja*: Model dievaluasi menggunakan metrik seperti *F1-score*, yang mengukur keseimbangan antara *precision* dan *recall*, untuk memastikan kualitas prediksi.

B. Penerapan Metode Random Forest

Pengolahan data pada penelitian ini dilakukan secara sistematis untuk memastikan data siap digunakan dalam pembangunan model prediksi menggunakan algoritma *Random Forest*. Tahapan dimulai dengan pembersihan data, meliputi penghapusan duplikasi, pengisian nilai kosong, serta eliminasi *outlier* dengan metode *Z-Score*. Variabel yang tidak relevan dengan tujuan penelitian, seperti nama siswa atau akun media sosial, juga dihapus untuk menyederhanakan analisis. Selanjutnya, data dengan banyak kategori ditransformasi menggunakan *MultiLabelBinarizer*, sementara variabel kategorikal lain dikonversi ke bentuk numerik agar dapat diproses dengan lebih efisien. Untuk menyamakan skala antarfitur digunakan *StandardScaler*, sedangkan ketidakseimbangan kelas diatasi dengan teknik *SMOTE* melalui penciptaan data sintesis pada kategori minoritas. Tahapan ini menghasilkan dataset yang bersih, seimbang, dan siap digunakan untuk pelatihan model.

Data yang telah diproses kemudian dibagi menjadi *training set* (80%) dan *testing set* (20%) dengan metode *stratified split* guna menjaga proporsi antara kategori “Kesepian Rendah” dan “Kesepian Tinggi”. Model *Random Forest* dibangun dengan menghasilkan sejumlah pohon keputusan yang dilatih secara acak pada *subset data* dan fitur berbeda. Prediksi akhir diperoleh melalui *majority voting*, yaitu penentuan label berdasarkan hasil terbanyak dari seluruh pohon keputusan. Mekanisme ini membuat *Random Forest* lebih stabil dan mampu mengurangi risiko *overfitting*. Proses tersebut digambarkan dengan jelas pada Gambar 1.



pencarian konfigurasi terbaik sehingga model dapat menghasilkan kinerja yang lebih efisien dan akurat. Evaluasi dilakukan dengan menggunakan data pengujian melalui metrik akurasi, *precision*, *recall*, dan *F1-score*, serta *confusion matrix* yang memberikan gambaran detail mengenai keberhasilan maupun kesalahan prediksi pada tiap kategori tingkat kesepian.

III. HASIL DAN PEMBAHASAN

A. Data Selection

Penelitian ini melibatkan 310 responden yang terbagi dalam tiga jenjang kelas, yakni Kelas 10 sebanyak 65 orang, Kelas 11 sebanyak 103 orang, dan Kelas 12 sebanyak 142 orang. Dari sisi jenis kelamin, mayoritas responden adalah laki-laki (66%), sedangkan perempuan hanya 34%. Berdasarkan usia, sebagian besar responden berada pada rentang 16–17 tahun, dengan dominasi usia 17 tahun (39%), disusul usia 16 tahun (28%), 18 tahun (24%), 15 tahun (8%), dan 19 tahun (2%). Data penelitian tersebut dikumpulkan melalui kuesioner yang dirancang untuk mengukur berbagai aspek penggunaan media sosial, meliputi pola aktivitas, jenis konten, interaksi sosial, ekspresi emosi, jaringan sosial, serta preferensi *platform*. Informasi ini digunakan sebagai dasar dalam membangun model prediksi tingkat kesepian responden. Hasil analisis menunjukkan bahwa sebagian besar responden menghabiskan waktu 1–3 jam per hari di media sosial (37,74%), diikuti penggunaan 3–5 jam (32,90%) dan lebih dari 5 jam (22,58%). Hanya sedikit responden yang menggunakannya kurang dari 1 jam (6,77%). Waktu penggunaan paling dominan adalah malam hari (65,81%), sedangkan penggunaan pada sore (20,65%), siang (10%), dan pagi (3,55%) relatif lebih rendah. Dari segi interaksi, mayoritas responden sering menggunakan fitur chat (44,52%), sementara 28,71% menggunakannya kadang-kadang. Aktivitas berbagi konten juga cukup umum, dengan 59,68% membuat postingan 1–2 kali per minggu dan 24,84% tidak pernah membuat postingan. *TikTok* (48%) dan *Instagram* (46%) menjadi platform utama, sedangkan *Facebook* hanya digunakan oleh 6% responden. Dalam hal keterlibatan sosial, sebagian besar responden memberikan komentar atau *like* hanya sesekali (40,32%), dan hanya 27,42% yang sering melakukannya. Penerimaan interaksi juga relatif rendah, dengan 40% responden jarang mendapat komentar atau *like*, serta 15% yang sering mendapatkannya. Pertumbuhan *follower* cenderung stagnan, di mana 59% jarang memperoleh *follower* baru dan hanya kurang dari 4% yang sering menambahkannya. Selain itu, kecenderungan berpindah *platform* cukup tinggi, dengan hampir 40% responden sering berpindah, 34% kadang-kadang, dan 24% jarang melakukannya.

Hasil pengukuran menggunakan *UCLA Loneliness Scale* menunjukkan bahwa mayoritas responden berada pada kategori kesepian rendah dengan jumlah 233 orang, sedangkan 77 responden termasuk dalam kategori kesepian tinggi. Temuan ini memperlihatkan bahwa sebagian besar responden cenderung memiliki tingkat keterhubungan sosial yang baik dan relatif jarang mengalami kesepian. Namun demikian, masih terdapat kelompok responden dengan tingkat kesepian

tinggi yang menunjukkan adanya pengalaman kesepian yang lebih intens, meskipun jumlahnya lebih sedikit.

B. Data Preprocessing

1) *Validasi Data*: Sebelum analisis dilakukan, data divalidasi untuk memastikan kualitas dan konsistensi jawaban responden. Validasi dilakukan melalui *Satisficing Theory* dengan mendeteksi pola jawaban tidak serius (misalnya *straight-lining* dan *low variation*) serta *Response Process Validity* untuk menilai konsistensi antarjawaban. Responden yang terbukti tidak konsisten atau kurang serius dikeluarkan agar tidak memengaruhi hasil analisis.

Tabel 1. Hasil Validasi Berdasarkan *Satisficing Theory*

Kategori Validasi	Jumlah Responden
Total Responden	309
Terindikasi <i>Straight-lining</i>	1
Terindikasi <i>Low Variation</i>	12
Total Indikasi Tidak Serius (gabungan)	12
Total Responden yang Dikeluarkan	12
Total Responden yang Layak Digunakan	297

2) *Pembersihan Data*: Tahapan ini meliputi penghapusan duplikasi, penanganan nilai kosong dengan metode rata-rata/median/modus, serta penghapusan *outlier* menggunakan metode *Z-Score*. Selain itu, variabel yang tidak relevan dengan tujuan penelitian seperti identitas responden dan metadata dihapus. Setelah proses ini, jumlah data berkurang dari 297 menjadi 260 entri yang lebih fokus dan siap dianalisis.

3) *Penanganan Data Multi-Label*: Beberapa pertanyaan kuesioner memiliki jawaban ganda (*multi-label*). Data tersebut diubah menggunakan *MultiLabelBinarizer* sehingga setiap kategori direpresentasikan dalam format biner (1 = muncul, 0 = tidak muncul). Transformasi ini menghasilkan struktur data yang lebih sistematis dan sesuai untuk model *machine learning*.

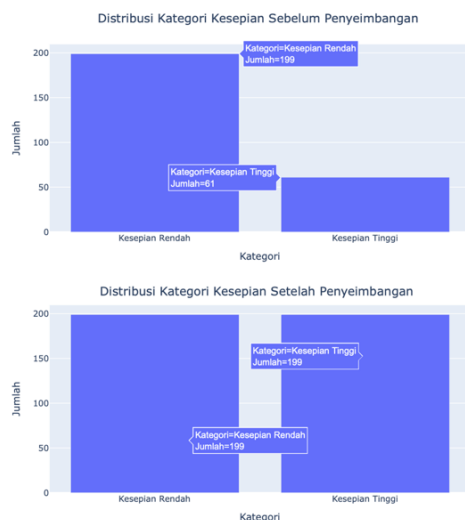
4) *Penyederhanaan Nama Fitur*: Beberapa nama kolom yang terlalu panjang disederhanakan agar lebih ringkas dan mudah digunakan dalam analisis maupun pemodelan. Misalnya, “Berapa lama waktu yang biasa Anda habiskan untuk menggunakan media sosial setiap hari?” diganti menjadi “Durasi harian penggunaan media sosial”, sedangkan “Berapa jumlah teman/follower Anda di media sosial?” menjadi “Jumlah follower”.

C. Data Transformation

1) *Transformasi Data*: Variabel kategorikal diubah menjadi numerik agar dapat diproses oleh algoritma machine learning. Contohnya, durasi penggunaan media sosial (<1 jam, 1–3 jam, 3–5 jam, >5 jam) dikodekan menjadi 0–3. Proses serupa dilakukan pada variabel lain sehingga dataset menjadi terstruktur secara numerik.

2) *Standarisasi Data*: Standarisasi dilakukan menggunakan *StandardScaler* agar setiap fitur memiliki rata-rata nol dan standar deviasi satu. Langkah ini memastikan konsistensi skala fitur sehingga meningkatkan kinerja model.

3) *Penyeimbangan Data dengan SMOTE*: Distribusi kelas yang tidak seimbang ditangani dengan *Synthetic Minority Over-sampling Technique (SMOTE)*. Jumlah sampel Kesenian Tinggi meningkat dari 61 menjadi 199, sehingga seimbang dengan kelas Kesenian Rendah.



Gambar 3. Hasil Penyeimbangan Data dengan SMOTE

D. Data Mining

Proses data mining pada penelitian ini mencakup tahapan pemodelan yang dilakukan secara sistematis melalui pemilihan algoritma, pelatihan, serta optimasi *hyperparameter*. Algoritma *Random Forest Classifier* dipilih karena keandalannya dalam mengolah data yang kompleks serta ketahanannya terhadap *overfitting* [12]. Algoritma ini bekerja dengan menggabungkan sejumlah *decision tree*, sehingga menghasilkan prediksi yang stabil dan akurat sekaligus memberikan informasi mengenai tingkat kepentingan fitur. Proses pelatihan dilakukan menggunakan data latih dengan variabel prediktor (X) dan target (y) berupa kategori tingkat kesenian, sementara evaluasi kinerja dilakukan pada data uji untuk menilai kemampuan generalisasi model. Selanjutnya, optimasi *hyperparameter* diterapkan menggunakan *Optuna* dengan pendekatan *Bayesian Optimization* yang divalidasi melalui *5-fold cross-validation*. Parameter yang dioptimasi meliputi *n_estimators*, *max_depth*, *min_samples_split*, *min_samples_leaf*, *max_features*, *bootstrap*, dan *criterion*. Dari hasil 10 percobaan, diperoleh kombinasi parameter terbaik seperti yang disajikan pada Tabel 2. Kombinasi tersebut terbukti mampu mencapai keseimbangan antara kompleksitas dan kemampuan generalisasi, sehingga menghasilkan model yang stabil dan akurat dalam memprediksi tingkat kesenian.

Tabel II. Parameter Metode *Random Forest*

Parameter	Nilai
<i>n_estimators</i>	2350
<i>max_depth</i>	20
<i>min_samples_split</i>	5

<i>min_samples_leaf</i>	1
<i>max_features</i>	'sqrt'
<i>bootstrap</i>	True
<i>criterion</i>	'entropy'

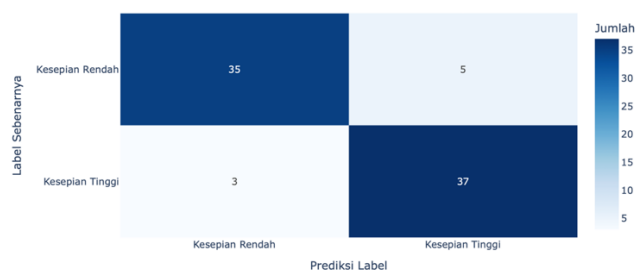
E. Data Evaluation

Kinerja model dievaluasi berdasarkan beberapa metrik, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*, serta dilengkapi dengan *confusion matrix* untuk menggambarkan distribusi prediksi dibandingkan dengan label aktual. Hasil evaluasi ini menjadi dasar penilaian konsistensi dan ketepatan klasifikasi model.

Tabel III. *Classification Report Random Forest*

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
Kesenian Rendah	0,92	0,88	0,90	40
Kesenian Tinggi	0,88	0,93	0,90	40
<i>Accuracy</i>			0,90	80
<i>Macro avg</i>	0,90	0,90	0,90	80
<i>Weighted avg</i>	0,90	0,90	0,90	80

Secara umum, model *Random Forest* mencapai tingkat akurasi sebesar 0,90, yang berarti 90% prediksi sesuai dengan label sebenarnya. Nilai *macro average* dan *weighted average* untuk *precision*, *recall*, serta *F1-score* juga konsisten pada angka 0,90. Hal ini menunjukkan bahwa model mampu mengklasifikasikan kedua kelas, Kesenian Rendah dan Kesenian Tinggi secara seimbang tanpa indikasi bias.



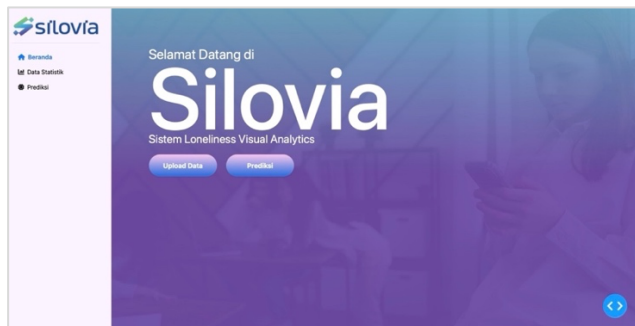
Gambar 4. *Confusion Matrix Random Forest*

F. Deployment

Perancangan prototipe dilakukan melalui *Figma* untuk menghasilkan antarmuka yang sederhana, responsif, dan mudah diakses, kemudian diimplementasikan menggunakan *Dash* yang mengintegrasikan model *machine learning* dengan sistem antarmuka. Aplikasi ini dirancang agar pengguna dapat memasukkan data secara langsung, memprosesnya melalui model, serta memperoleh hasil prediksi secara *real-time*. Selain itu, basis data *MySQL* digunakan untuk menyimpan hasil kuesioner penelitian yang selanjutnya divisualisasikan pada *dashboard*, sehingga memudahkan proses analisis data.

Secara keseluruhan, *prototipe* aplikasi terdiri atas tiga fitur utama.

- 1) *Fitur Pengujian Model*: berfungsi untuk melatih model menggunakan dataset yang diunggah serta menampilkan performa dan visualisasi hasilnya.
- 2) *Fitur Prediksi*: memungkinkan pengguna memperoleh hasil klasifikasi berdasarkan data profil media sosial.
- 3) *Fitur Data Statistik*: menyajikan visualisasi demografi responden dan pola penggunaan media sosial.



Gambar 5. Prototipe Prediksi Kesepian

G. Pembahasan

Berdasarkan hasil penelitian, algoritma *Random Forest* terbukti mampu memprediksi tingkat kesepian siswa SMK secara efektif dengan memanfaatkan data profil media sosial yang telah dilabeli menggunakan kuesioner *UCLA Loneliness Scale V.3*. Model yang dioptimasi dengan *Optuna* menunjukkan performa terbaik dengan nilai akurasi, presisi, *recall*, dan *F1-score* sebesar 0,90, menandakan kemampuan klasifikasi yang tinggi dan seimbang pada kedua kategori kesepian. Hasil prediksi model juga memiliki tingkat kesesuaian yang tinggi dengan data label dari kuesioner, sehingga memperkuat validitas model tidak hanya secara matematis, tetapi juga dalam konteks psikologis yang telah tervalidasi. Hal ini menunjukkan bahwa *Random Forest* dapat menjadi alat bantu yang andal dan efisien dalam mendeteksi kondisi kesepian siswa.

Secara keseluruhan, penelitian ini menunjukkan bahwa pendekatan berbasis data dan teknologi dapat dimanfaatkan secara nyata dalam dunia pendidikan, khususnya dalam membantu sekolah mendeteksi masalah kesejahteraan psikologis siswa sejak dini. Dengan memanfaatkan data yang berasal dari aktivitas media sosial siswa, sekolah dapat melakukan pemantauan secara lebih cepat dan tidak mengganggu aktivitas belajar. Temuan ini membuka peluang baru untuk mengembangkan sistem pemantauan psikologis yang bersifat preventif, berbasis teknologi, dan responsif terhadap kebutuhan siswa. Dengan begitu, pendidikan tidak hanya berfokus pada aspek kognitif semata, tetapi juga memberikan perhatian pada kesejahteraan mental peserta didik.

IV. KESIMPULAN

Penelitian ini bertujuan untuk membangun sebuah model prediksi yang mampu mengidentifikasi tingkat kesepian pada siswa Generasi Z di SMK Tunas Media dengan memanfaatkan

data profil media sosial menggunakan algoritma *Random Forest*. Hasil pengujian menunjukkan bahwa *Random Forest* yang dioptimalkan dengan *Optuna* memberikan performa terbaik dengan nilai *accuracy*, *precision*, *recall*, dan *F1-score* yang sama tinggi, yaitu 0,90.

Pengolahan data dilakukan melalui serangkaian tahapan, mulai dari pembersihan data, penanganan data dengan kategori ganda (*multi-label*), penyederhanaan nama-nama fitur, transformasi data kategorikal ke bentuk numerik, standarisasi data untuk menyeragamkan skala antar fitur, hingga penyeimbangan data agar distribusi antar kelas menjadi lebih proporsional. Data yang digunakan merupakan hasil dari kuesioner *UCLA Loneliness Scale V.3* yang telah dikategorikan menjadi dua tingkat kesepian, yaitu rendah dan tinggi.

Secara umum, penelitian ini menunjukkan bahwa data dari media sosial dapat dimanfaatkan secara efektif untuk membangun model prediksi yang akurat terkait kondisi psikologis siswa, khususnya perasaan kesepian. Selain itu, aplikasi berbasis web yang dikembangkan dalam penelitian ini mampu mengintegrasikan model prediktif ke dalam sistem yang interaktif dan mudah digunakan. Temuan ini tidak hanya menjawab rumusan masalah dalam penelitian, tetapi juga memberikan manfaat praktis bagi sekolah dalam upaya deteksi dini dan intervensi terhadap masalah kesejahteraan mental siswa secara lebih tepat sasaran dan berbasis teknologi.

V. SARAN

Berdasarkan hasil yang telah diperoleh, beberapa saran berikut dapat dijadikan pertimbangan untuk pengembangan penelitian di masa yang akan datang:

- 1) Melihat keterbatasan penelitian yang hanya dilakukan pada satu sekolah dengan jumlah responden terbatas, disarankan agar penelitian di masa mendatang dapat mencakup lebih banyak sekolah atau wilayah lain. Hal ini penting untuk mendapatkan hasil yang lebih beragam dan representatif.
- 2) Agar analisis menjadi lebih komprehensif, sebaiknya ditambahkan variabel psikologis lain yang juga relevan, seperti tingkat stres, kecemasan, atau rasa percaya diri siswa. Penambahan variabel ini akan memperluas cakupan dan memberikan gambaran yang lebih menyeluruh terhadap kondisi psikologis siswa.
- 3) Dari sisi pengembangan aplikasi, masih terdapat ruang untuk penyempurnaan, misalnya dengan menambahkan fitur lanjutan berupa rekomendasi atau tindak lanjut berdasarkan hasil prediksi. Fitur tersebut juga bisa diintegrasikan langsung dengan layanan konseling di sekolah, sehingga hasil prediksi dapat segera ditindaklanjuti secara profesional.
- 4) Pada aspek teknis pemodelan, penyesuaian parameter model secara lebih optimal masih sangat mungkin dilakukan untuk meningkatkan kualitas prediksi. Teknik optimasi *hyperparameter* seperti *Grid Search* atau *Random Search* dapat diterapkan untuk menemukan konfigurasi terbaik dari masing-masing algoritma yang digunakan.

REFERENSI

- [1] A. Ahdiat, "Media Sosial Favorit Gen Z dan Milenial Indonesia," Databoks. Diakses: 30 November 2024. [Daring]. Tersedia pada: <https://databoks.katadata.co.id/infografik/2024/09/24/media-sosial-favorit-gen-z-dan-milenial-indonesia>
- [2] D. Russell, L. A. Peplau, dan C. E. Cutrona, "The Revised UCLA Loneliness Scale: Concurrent and Discriminant Validity Evidence," Sep 1980.
- [3] A. Asrori, "Generasi Z Dinobatkan Jadi Generasi Paling Kesepian, Berikut Efek Bahaya dan 3 Penyebab di Baliknyanya," Jawa Pos. Diakses: 2 Desember 2024. [Daring]. Tersedia pada: <https://www.jawapos.com/lifestyle/014629123/generasi-z-dinobatkan-jadi-generasi-paling-kesepian-berikut-efek-bahaya-dan-3-penyebab-di-baliknyanya>
- [4] B. d'Hombres dan C. Gentile, "Social Media Use and Loneliness," 2024, hlm. 93–115. doi: 10.1007/978-3-031-66582-0_5.
- [5] B. Prawira, S. Magdalena, M. Jenifer Siandita, H. Rachmadiani Sukma, dan A. Liem, "Seri Laporan ke-1 Laporan Perilaku Penggunaan Layanan Kesehatan Mental di Indonesia - Hasil Awal," Sep 2021. [Daring]. Tersedia pada: www.intothelightid.org.
- [6] J. Zhang *dkk.*, "Predicting the Risk of Loneliness in Children and Adolescents: A Machine Learning Study," *Behavioral Sciences*, vol. 14, no. 10, Okt 2024, doi: 10.3390/bs14100947.
- [7] E. Ahmed *dkk.*, "Predicting Loneliness through Digital Footprints on Google and YouTube," *Electronics (Switzerland)*, vol. 12, no. 23, Des 2023, doi: 10.3390/electronics12234821.
- [8] M. Panayiotou, J. C. Badcock, M. H. Lim, M. J. Banissy, dan P. Qualter, "Measuring Loneliness in Different Age Groups: The Measurement Invariance of the UCLA Loneliness Scale," *Assessment*, vol. 30, no. 5, hlm. 1688–1715, Jul 2023, doi: 10.1177/10731911221119533.
- [9] R. Bottaro, G. D. Valenti, dan P. Faraci, "Assessment of an Epidemic Urgency: Psychometric Evidence for the UCLA Loneliness Scale," *Psychol Res Behav Manag*, vol. 16, hlm. 2843–2855, 2023, doi: 10.2147/PRBM.S406523.
- [10] C. J. Gosling, R. Colle, A. Cartigny, F. Jollant, E. Corruble, dan A. Frajerman, "Measuring loneliness: a head-to-head psychometric comparison of the 3- and 20-item UCLA Loneliness Scales," *Psychol Med*, Okt 2024, doi: 10.1017/S0033291724002083.
- [11] Utkarsh, "Knowledge Discovery in Databases." Diakses: 4 Januari 2025. [Daring]. Tersedia pada: <https://www.scaler.com/topics/data-mining-tutorial/kdd-in-data-mining/>
- [12] A. S. Bello-Valle, A. Martínez-Rebollar, W. Sánchez, dan H. Estrada-Esquivel, "A Predictive Model for Automatic Detection of Loneliness and Social Isolation using Machine Learning," *Computacion y Sistemas*, vol. 26, no. 1, hlm. 113–124, 2022, doi: 10.13053/CyS-26-1-4157.