



Sentimen Analisis Kesehatan Mental *Anxiety* dengan Metode *Decision Tree* Menggunakan *Software Orange*

Eva Fauziah¹, Makhsun²

^{1,2)} Program Studi Teknik Informatika S-2, Universitas Pamulang

Email: ¹evafauziah112234@gmail.com, ²dosen00345@unpam.ac.id

ABSTRACT

Mental health, particularly anxiety disorders, has become a global concern as the prevalence of mental health issues increases worldwide. Anxiety can impact an individual's quality of life and productivity, making it crucial to analyze and detect its symptoms accurately. This study aims to apply the decision tree method to analyze anxiety sentiments embedded in texts from various sources, such as mental health forums and social media. The sample data used is sourced from Kaggle, consisting of 610 English text samples taken from Reddit and Twitter posts, labeled as Anxiety and Normal. The decision tree method was chosen due to its simplicity and effectiveness in classifying data based on specific patterns. The Orange software was used to build the classification model because of its ease of use and visualization capabilities. Based on the analysis of text data, consisting of 502 training data and 108 testing data, the evaluation matrix yielded the following results: accuracy 98,2%, precision 98,3%, and recall 98,2%. Although the model's identification is limited to expressions of anxiety in the text, it showed high prediction accuracy on the test data, with 100% prediction accuracy for the anxiety class and 100% for the normal class. This study demonstrates that the decision tree model is effective in identifying anxiety patterns in text.

Keywords: Mental health, anxiety, sentiment analysis, decision tree, Orange

ABSTRAK

Kesehatan mental, khususnya gangguan kecemasan (anxiety), semakin menjadi perhatian global seiring dengan meningkatnya angka gangguan mental di berbagai belahan dunia. Kecemasan dapat mempengaruhi kualitas hidup individu dan produktivitasnya, sehingga penting untuk dapat menganalisis dan mendeteksi gejalanya secara tepat. Penelitian ini bertujuan untuk menerapkan metode decision tree dalam menganalisis sentimen kecemasan yang terkandung dalam teks dari berbagai sumber seperti forum kesehatan mental dan media sosial. Data sampel yang digunakan bersumber dari Kaagle, jumlahnya 610 data teks dalam Bahasa Inggris yang diambil dari postingan sosial media Reddit dan Twitter dengan label kelas Anxiety dan Normal. Metode decision tree dipilih karena kemampuannya yang sederhana namun efektif dalam mengklasifikasikan data berdasarkan pola-pola tertentu. Software Orange digunakan untuk membangun model klasifikasi ini, karena kemudahan penggunaan dan visualisasi yang ditawarkannya. Berdasarkan hasil analisis pada data teks yang terdiri dari 502 data pelatihan dan 108 data pengujian, diperoleh matriks evaluasi sebagai berikut: akurasi 98,2%, presisi 98,3%, dan recall 98,2%. Meski identifikasi model terbatas pada ekspresi kecemasan yang muncul pada teks, akan tetapi model menunjukkan ketepatan prediksi yang tinggi terhadap data pengujian, dengan tingkat prediksi 100% untuk kelas anxiety dan 100% untuk kelas normal. Penelitian ini menunjukkan bahwa model decision tree efektif dalam mengidentifikasi pola kecemasan dalam teks.

Kata kunci: Kesehatan Mental, Kecemasan, Analisis Sentimen, *Decision Tree*, *Orange*

1. PENDAHULUAN

Kesehatan mental menjadi salah satu isu global yang semakin mendapat perhatian serius, terlebih dengan meningkatnya angka gangguan kesehatan mental di berbagai belahan dunia. Di antara gangguan kesehatan mental yang paling banyak dilaporkan, kecemasan (*anxiety*) merupakan salah satu kondisi yang paling sering dialami oleh individu, baik di kalangan remaja, dewasa, maupun lansia. Gangguan kecemasan adalah kondisi ketakutan berlebihan yang disertai dengan perilaku tidak nyaman dan sulit dikendalikan terhadap hal-hal yang tidak pasti [1]. Kecemasan dapat merugikan dan mengancam kesehatan baik fisik maupun psikis karena gangguan mental anxiety menimbulkan perasaan resah, ragu, tegang, takut, bimbang, dan curiga terhadap sesuatu yang belum pasti terjadi [2]. Gangguan kecemasan ini tidak hanya berdampak pada kesejahteraan individu, tetapi juga mempengaruhi produktivitas, hubungan sosial, dan kualitas hidup secara keseluruhan. Untuk memahami dan menangani kecemasan secara efektif, penting untuk dapat mendekripsi dan menganalisis gejala-gejalanya secara tepat.

Analisis sentimen adalah teknik otomatis untuk memahami, mengekstrak, dan mengolah data tekstual guna memperoleh wawasan atau informasi yang relevan dan bermanfaat [3]. Melalui sentimen analisis, data teks seperti postingan media sosial, ulasan pengguna, atau wawancara dapat dianalisis untuk mengidentifikasi tanda-tanda atau pola-pola yang berkaitan dengan kecemasan. Namun, salah satu tantangan utama dalam analisis ini adalah ketepatan dan akurasi dalam mengklasifikasikan tingkat kecemasan individu.

Decision tree adalah pendekatan yang cukup sederhana untuk dimengerti dan diterapkan, dengan kemampuan untuk mengklasifikasikan data berdasarkan atribut-atribut tertentu yang teridentifikasi dalam dataset. Decision Tree mempartisi ruang pencarian menjadi masalah-masalah kecil dan mengubah data tabel menjadi model pohon terstruktur untuk mempermudah pengambilan keputusan [4]. Dalam konteks analisis sentimen untuk kesehatan mental, decision tree dapat digunakan untuk mengidentifikasi pola-pola kecemasan dari teks yang dihasilkan oleh pengguna atau pasien.

Software *Orange* menyediakan berbagai widget untuk mengolah data informatif, termasuk data dari isi status dan tanggapan akun Twitter [5]. *Orange* memungkinkan pengguna untuk membangun model *Decision Tree* tanpa memerlukan keterampilan

pemrograman yang mendalam, sehingga menjadi pilihan yang tepat untuk memfasilitasi penelitian ini. Meskipun berbagai metode telah diterapkan dalam analisis sentimen terkait kesehatan mental, pemanfaatan *Decision Tree* dengan perangkat lunak *Orange* masih terbilang terbatas dalam literatur yang ada.

Pada penelitian sebelumnya, seperti yang dilakukan oleh Firmansyah dkk [6], telah mengembangkan model pembelajaran mesin untuk mengkaji kesehatan mental karyawan di lingkungan kerja, namun belum banyak yang mengkaji penerapan metode *Decision Tree* dalam analisis kecemasan secara spesifik. Selanjutnya, penelitian oleh Muhammad Daffa Alfahreza dkk [7], menganalisis sentimen kesehatan mental terhadap generasi Z terkait pengaruh jam kerja, namun metode yang digunakan lebih sering mengandalkan algoritma yang lebih kompleks, seperti *Support Vector Machines (SVM)* dan *Random Forests*. Kelebihan dari *Decision Tree* adalah interpretabilitasnya yang lebih tinggi dibandingkan dengan metode-metode lain, memberikan gambaran yang jelas mengenai bagaimana keputusan dibuat dalam mengklasifikasikan kecemasan.

Selain itu, penelitian oleh Yunis Femilia Nugraini dkk [8], menunjukkan bahwa banyak pendekatan dalam analisis sentimen kesehatan mental berfokus pada deteksi hanya berdasarkan kondisi emosional seperti sedih dan bahagia, sementara kecemasan seringkali membutuhkan pemahaman yang lebih mendalam terkait dengan konteks psikologis dan faktor-faktor individual. Untuk itu, gap dalam penelitian ini adalah penerapan metode yang lebih dapat diinterpretasi dan berfokus pada aspek-aspek yang lebih spesifik dalam kecemasan. *Decision tree* diharapkan dapat memberikan pendekatan yang lebih terstruktur dan transparan dalam mengidentifikasi kecemasan dari teks yang dihasilkan oleh individu.

Adapun tujuan penelitian ini agar dapat menerapkan metode *Decision Tree* dalam menganalisis sentimen terkait kecemasan pada teks yang dikumpulkan dari berbagai sumber. Penelitian ini akan menggunakan software *Orange* untuk membangun model klasifikasi yang dapat mengidentifikasi kecemasan berdasarkan analisis sentimen. Dengan demikian, penelitian ini tidak hanya berkontribusi pada pemahaman sentimen analisis untuk kesehatan mental, tetapi juga memperkenalkan pendekatan yang lebih mudah diakses dan dapat diinterpretasi oleh praktisi dan peneliti dalam bidang kesehatan mental.

2. METODE

Dalam rangka mencapai tujuan penelitian, maka dilakukan serangkaian tahapan yang terstruktur, yang meliputi pengumpulan data, persiapan data, pembuatan model, pengujian dan evaluasi model. Setiap tahapan dirancang untuk memastikan kualitas dan ketepatan hasil yang diperoleh.

2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data teks yang mengandung gejala kecemasan yang teridentifikasi dalam konteks kesehatan mental. Data yang digunakan bersumber dari Kaggle dimana sumber data utama yang digunakan adalah komentar dari media sosial, postingan Reddit, postingan Twitter dan lainnya yang melibatkan responden dengan keluhan terkait kecemasan. Adapun sumber link: <https://www.kaggle.com/datasets/suchintikasarkar/sentiment-analysis-for-mental-health>

2.2. Persiapan Data

Setelah pengumpulan data, tahap berikutnya adalah persiapan data. Tahap persiapan data bertujuan mengubah data mentah yang tidak teratur menjadi data yang lebih terorganisir dan terstruktur [9]. Diantaranya meliputi pembersihan dan transformasi data agar siap untuk dianalisis. Data teks yang tidak relevan atau tidak lengkap akan dibuang, sementara teks yang dianggap relevan akan dianalisis lebih lanjut. Kemudian tahap Tokenisasi yaitu pemecahan teks menjadi bagian-bagian kata atau frasa dan Filtering dengan penghapusan stopwords atau menghapus kata-kata generik yang tidak menyampaikan informasi penting terhadap analisis sentimen. Kemudian, data akan diubah menjadi format numerik yang dapat dipahami oleh algoritma decision tree, melalui teknik representasi seperti *bag of words* atau *TF-IDF (Term Frequency-Inverse Document Frequency)*.

2.3. Pembuatan Model Data Mining

Data mining adalah proses menganalisis data, baik terstruktur maupun tidak terstruktur, untuk menemukan pola, hubungan, dan informasi penting lainnya [10]. Penelitian ini menggunakan model data mining berupa decision tree yang dibangun menggunakan perangkat lunak *Orange*. Dataset yang digunakan terdiri dari 610 data teks dalam Bahasa Inggris dengan label kelas Anxiety dan Normal. Analisis data dilakukan

dalam tiga percobaan pelatihan, dengan pembagian data sebagai berikut: 50% data pelatihan dan 50% data pengujian untuk percobaan pertama, 75% data pelatihan dan 25% data pengujian untuk percobaan kedua, serta 82% data pelatihan dan 18% data pengujian untuk percobaan ketiga seperti yang ditunjukkan di bawah ini:

Tabel 2.1 Data Sampel

Percobaan	Data Pelatihan (%)	Jumlah Data Pelatihan	Data Pengujian (%)	Jumlah Data Pengujian
Pelatihan 1	50%	305	50%	305
Pelatihan 2	75%	457	25%	153
Pelatihan 3	82%	502	18%	108

Serangkaian percobaan pelatihan data dilakukan dengan tujuan untuk mengidentifikasi model klasifikasi terbaik dalam memprediksi kondisi anxiety dan normal. Hasil evaluasi kinerja model data mining, yang diukur melalui perangkat lunak Orange, dari ketiga percobaan pelatihan tersebut disajikan di bawah ini:

Tabel 2.2 Peluang Data Sampel

Pelatihan ke-	Data Pelatihan	Data Pengujian	Akurasi Model	Akurasi Data Baru
1	305	305	97%	98,7%
2	457	153	98,2%	99,3%
3	502	108	98,2%	100%

Berdasarkan data yang tertera di tabel, dapat disimpulkan bahwa model klasifikasi terbaik diperoleh dengan menggunakan data pelatihan sebanyak 502 sampel dan data pelatihan sebanyak 108 sampel. Hal ini dikarenakan evaluasi model tersebut menghasilkan akurasi sebesar 98,2% dan mampu memprediksi data baru dengan akurasi mencapai 100%.

2.4. Rumus Entropy

Entropi dalam pohon keputusan mengukur tingkat ketidakpastian data. Semakin tinggi entropi, semakin acak data tersebut; semakin rendah entropi, semakin terstruktur data tersebut. Rumus Entropy sebagai berikut:

$$E(S) = - \sum_{i=1}^n p_i * \log_2 p_i \quad (1)$$

$E(S)$ adalah entropi dari set data S, p_i adalah probabilitas relatif dari kelas i dalam set data S, n adalah jumlah kelas yang ada dalam data, \log_2 adalah logaritma basis 2, yang digunakan untuk menghitung entropi dalam satuan bit.

Information Gain (IG) mengukur pengurangan ketidakpastian (entropi) setelah data dibagi berdasarkan fitur tertentu. Rumus IG sebagai berikut:

$$IG(S, A) = E(S) - \sum_{v \in values(A)} \frac{|S_v|}{|S|} E(S_v) \quad (2)$$

S adalah set data yang ingin dibagi, A adalah atribut atau fitur yang dipertimbangkan untuk pembagian, $values(A)$ adalah semua nilai yang mungkin dari fitur A, S_v adalah subset data berdasarkan nilai v dari fitur A, $|S|$ adalah ukuran set data S.

2.5. Evaluasi

Tahap evaluasi dan validasi bertujuan untuk memastikan bahwa model yang dibangun dapat diterapkan dalam konteks yang lebih luas dan memberikan hasil yang valid dan reliabel. *Confusion matrix* adalah tabel untuk mengevaluasi performa model klasifikasi dengan membandingkan prediksi model dengan label sebenarnya. Tabel ini menunjukkan jumlah data yang diklasifikasikan dengan benar atau salah, membantu memahami kesalahan model dalam memprediksi kelas. Performa dari *confusion matrix* terdiri dari *Accuracy*, *Precision*, *Recall* dan *F1 Score* dengan komponen dasar seperti:

Tabel 2.3 Confusion Matrix

Nilai Prediksi	Nilai Aktual	
	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

Penjelasan tabel bahwa untuk *True Positive (TP)* maka jumlah data yang benar-benar positif dan diprediksi sebagai positif oleh model, *True Negative (TN)* maka jumlah data yang benar-benar negatif dan diprediksi sebagai negatif oleh model, *False Positive (FP)* maka jumlah data yang sebenarnya negatif, tetapi diprediksi sebagai positif oleh model, *False Negative (FN)* maka jumlah data yang sebenarnya positif, tetapi diprediksi sebagai negatif oleh model. Komponen dasar ini digunakan dalam menentukan performa *confusion matrix* dengan rumus seperti berikut:

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \times 100 \% \quad (3)$$

$$Precision = \frac{TP}{(TP + FP)} \times 100 \% \quad (4)$$

$$Recall = \frac{TP}{(TP + FN)} \times 100 \% \quad (5)$$

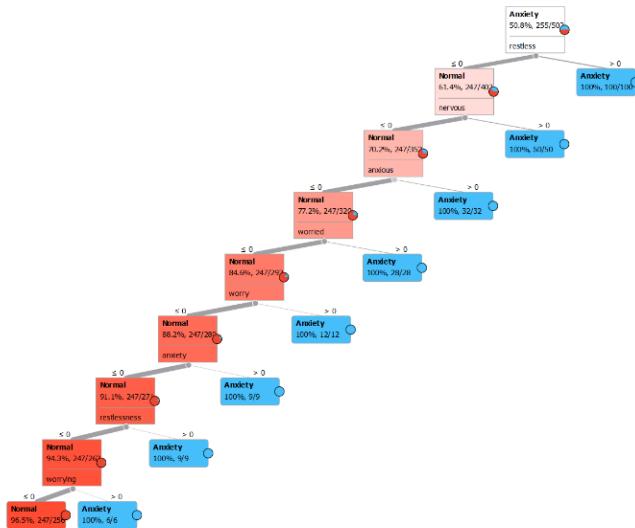
$$F1 Score = 2 * (Precision * Recall) / (Precision + Recall) \quad (6)$$

Pada konteks evaluasi model, *Accuracy* menunjukkan persentase prediksi yang benar dari seluruh prediksi yang dilakukan. *Precision* menunjukkan banyaknya prediksi positif yang benar-benar positif, di antara semua prediksi positif yang dibuat oleh model. *Recall* menunjukkan banyaknya data positif yang berhasil diprediksi dengan benar oleh model, di antara seluruh data yang seharusnya positif dan *F1 Score* menunjukkan rata-rata harmonis antara precision dan recall.

3. HASIL DAN PEMBAHASAN

3.1. Hierarki Decision Tree

Setelah melalui tahapan prapemrosesan dan pengolahan data teks, sistem melakukan seleksi fitur yang menghasilkan atribut-atribut yang berperan dalam pembentukan *Decision Tree*. Pada tahap ini, kata kunci sebagai atribut yang relevan dapat diidentifikasi melalui *Tree Viewer* dengan terlebih dahulu menerapkan metode *Decision Tree* pada data pelatihan untuk membangun model klasifikasi. Berikut adalah hierarki pohon keputusan yang menampilkan fitur kata kunci *Restless*, *Nervous*, *Anxious*, *Worried*, *Worry*, *Anxiety*, *Restlessness*, dan *Worrying*.



Gambar 3.1 Tree Viewer

Deskripsi menurut gambar, node akar atau root node adalah titik awal, dalam hal ini dimulai dengan pertanyaan "restless" (gelisah). Dari node akar, terdapat dua cabang yang mewakili jawaban "ya" atau "tidak". Jika seseorang menjawab "ya", maka proses akan berlanjut ke node berikutnya di cabang kanan. Sebaliknya, jika seseorang menjawab "tidak", maka proses akan berpindah ke node berikutnya di cabang kiri.

Setiap cabang akan membawa ke node anak baru, di mana pertanyaan lanjutan akan diajukan. Proses ini berlanjut hingga mencapai node terakhir, yang disebut daun (leaf). Daun ini mewakili hasil akhir klasifikasi, yaitu "Anxiety" (kecemasan) atau "Normal".

Angka persentase pada setiap daun menunjukkan tingkat keyakinan model terhadap klasifikasi tersebut. Sebagai contoh, jika sebuah daun bertuliskan "Anxiety 100%, 100/100", ini berarti model sepenuhnya yakin bahwa individu yang mencapai daun tersebut mengalami kecemasan berdasarkan data yang telah dilatih.

3.2. Hasil Analisis Model Decision Tree

Untuk hasil evaluasi keseluruhan dalam pembuatan model data mining menggunakan software orange dapat ditelaah seperti berikut:

Tabel 3.1 Hasil Analisis Model

Metrik	Nilai
AUC	0,973
CA (Akurasi)	0,982
F1 Score	0,982
Presisi	0,983
Recall	0,982
MCC	0,965

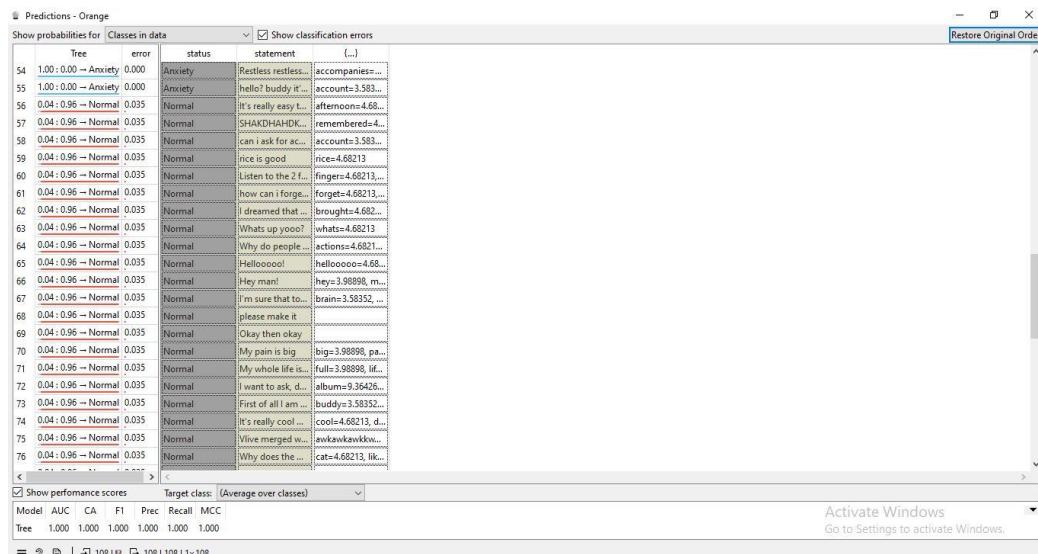
Berdasarkan hasil yang tertera pada tabel dengan menggunakan data pelatihan berjumlah 502 dan data pengujian berjumlah 108 yaitu, untuk AUC 0,973 menunjukkan bahwa model pohon keputusan untuk kelas positif dan negatif dapat dibedakan dengan sangat baik, CA 0,982 menunjukkan model benar dalam memprediksi kelas sebesar 98,2% dari semua data, F1-Score 0,982 menunjukkan keseimbangan mengukur yang baik antara precision dan recall, Precision 0,983 menunjukkan dari semua prediksi positif yang diberikan model 98,3% benar-benar positif, Recall 0,982 model berhasil mengidentifikasi 98,2% dari semua contoh positif yang sebenarnya ada dalam data, MCC 0,965 menunjukkan nilai yang mendekati 1 menunjukkan korelasi yang kuat antara prediksi dan label aktual.

Tabel 3.2 Hasil Confusion Matrix Data Pelatihan

Aktual/Prediksi	Anxiety	Normal	F1-Score	Presisi	Recall
Anxiety	246 (TP)	9 (FN)	0,982	1	0,965
Normal	0 (FP)	247 (TN)	0,982	0,965	1

Pada *confusion matrix*, untuk angka-angka yang diberi warna pada diagonal dari kiri atas ke kanan bawah menunjukkan jumlah data diklasifikasikan dengan benar dan

angka-angka di luar diagonal menunjukkan jumlah data yang salah diklasifikasikan. Adanya 9 *false negative*, artinya terdapat 9 data sempel dengan kelas anxiety yang teridentifikasi kelas normal dengan tingkat error 0,035. Hasil tabel evaluasi (*F1 Score, Presisi, Recall*) masing-masing kelas menunjukkan adanya hubungan tingkat kesalahan prediksi yang mempengaruhi ketepatan prediksi untuk data baru, analisis prediksi dapat dilihat seperti dibawah ini:



Gambar 3.1 Hasil Prediksi Model

Pada evaluasi data baru sebagai data pengujian, teridentifikasi tingkat kesalahan model dalam memprediksi sebesar 0,035. Namun model masih tepat memprediksi data pengujian kedalam kelas normal sebesar 0,965. Adapun sistematisnya, model bekerja dengan membuat serangkaian keputusan berdasarkan fitur-fitur yang ada dalam teks. Untuk menentukan sebuah teks mengandung sebuah teks mengandung ekspresi kecemasan atau tidak, model akan memeriksa keberadaan kata-kata berdasarkan kata kunci. Nilai antara 0 dan 1 yang menunjukkan tingkat kepercayaan model terhadap prediksinya. Semakin mendekati 1, semakin yakin model bahwa prediksinya benar. Adapun hasil prediksi model terhadap data baru sebagai berikut:

Tabel 3.3 Hasil Prediksi Model

Aktual/Prediksi	Anxiety	Normal	Jumlah
Anxiety	55 (TP)	0 (FN)	55
Normal	0 (FP)	53 (TN)	53

Jumlah	55	53	108
--------	----	----	-----

Hasil evaluasi dari data pengujian menunjukkan 55 data yang sebenarnya "Anxiety" berhasil diprediksi sebagai "Anxiety" dan 53 data yang sebenarnya "Normal" berhasil diprediksi sebagai "Normal".

4. KESIMPULAN

Penelitian ini menyimpulkan bahwa metode decision tree mampu mengidentifikasi pola kecemasan dalam teks menggunakan kata kunci seperti *Restless*, *Nervous*, *Anxious*, *Worried*, *Worry*, *Anxiety*, *Restlessness*, dan *Worrying* sebagai atribut utama. Model klasifikasi yang dihasilkan menunjukkan kinerja tinggi dengan skor evaluasi AUC 0.973, CA 0.982, F1-Score 0.982, Precision 0.983, Recall 0.982, dan MCC 0.965. Kinerja ini memungkinkan model mengklasifikasikan data baru dengan akurasi tinggi berdasarkan kata-kata yang mencerminkan kecemasan. Dengan demikian, metode decision tree terbukti efektif untuk mengidentifikasi teks yang mengandung ekspresi kecemasan secara akurat.

5. DAFTAR PUSTAKA

- [1] G. Zahra, N. Fadhilah, R. A. Saputra, and A. H. Wibowo, "Deteksi Tingkat Gangguan Kecemasan Menggunakan Metode Random Forest," *J. Fak. Tek. UMT*, vol. 13, no. 1, pp. 38–47, 2024, [Online]. Available: <http://jurnal.umt.ac.id/index.php/jt/index>
- [2] N. A. Pasa, Y. Maulita, and I. G. Prahmana, "Penggunaan Metode Rough Set pada Tingkat Kecemasan (Anxietas) Mahasiswa dalam Menyusun Tugas Akhir," no. 4, 2024.
- [3] A. Yahyadi and F. Latifah, "Analisis Sentimen Twitter TerhadapKebijakan Ppkm Di Tengah Pandemi Covid-19Menggunakan Mode Lstm," *J. Inf. Syst. Applied, Manag. Account. Res.*, vol. 6, no. 2, pp. 464–470, 2022, doi: 10.5236/jisamar.v6i2.791.
- [4] M. A. Ramadhan and M. I. Wahyudin, "Analisis Sentimen Mengenai Keberhasilan Indonesia di Ajang Thomas Cup 2020 (Studi Kasus Media Sosial Twitter)

- Menggunakan Metode Naïve Bayes dan Decision Tree,” *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 6, no. 4, pp. 505–511, 2022, doi: 10.35870/jtik.v6i4.560.
- [5] O. B. Islamuddin and I. Yuadi, “Analisis Text Clustering Pada Data Mining Menggunakan Metode K- Nearest Neighbor (Knn) Dan Decision Tree,” *J. Khatulistiwa Inform.*, vol. 11, no. 2, pp. 128–134, 2023.
- [6] F. Firmansyah and A. Yulianto, “Pemodelan Pembelajaran Mesin untuk Prediksi Kesehatan Mental di Tempat Kerja,” *J. Minfo Polgan*, vol. 13, no. 1, pp. 397–407, 2024, doi: 10.33395/jmp.v13i1.13674.
- [7] Muhammad Daffa Al Fahreza, Ardytha Luthfiarta, Muhammad Rafid, and Michael Indrawan, “Analisis Sentimen: Pengaruh Jam Kerja Terhadap Kesehatan Mental Generasi Z,” *J. Appl. Comput. Sci. Technol.*, vol. 5, no. 1, pp. 16–25, 2024, doi: 10.52158/jacost.v5i1.715.
- [8] Y. F. Nugraini, R. Rohmat Saedudin, and R. Andreswari, “Implementasi Data Mining Dalam Kasus Mental Health Pada Sosial Media Twitter Menggunakan Metode Naive Bayes,” *e-Proceeding Eng.*, vol. 8, no. 5, pp. 9260–9265, 2021.
- [9] S. M. Tambunan, Y. Nataliani, and E. S. Lestari, “Perbandingan Klasifikasi dengan Pendekatan Pembelajaran Mesin untuk Mengidentifikasi Tweet Hoaks di Media Sosial Twitter,” *J. Edukasi dan Penelit. Inform.*, vol. 7, no. 2, p. 112, 2021, doi: 10.26418/jp.v7i2.47232.
- [10] M. R. Firdaus, N. Rahatingsih, and R. D. Dana, “Analisis Sentimen Aplikasi Shopee di Goole Play Store Menggunakan Klasifikasi Algoritma Naïve Bayes,” *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 6, no. 1, pp. 228–237, 2024.