



Klasifikasi Multi Label untuk Deteksi Keseimbangan Emosi Pengguna Media Sosial Menggunakan K-Fold Cross Validation

Titik Misriati¹, Riska Aryanti^{1,*}, Asriyani Sagiyanto², Muhamad Fachri¹, Arya Ramadhani¹

¹Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Jakarta

Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta, Indonesia

²Fakultas Komunikasi dan Bahasa, Program Studi, Universitas Bina Sarana Informatika, Jakarta

Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta, Indonesia

Email: ¹titik.tmi@bsi.ac.id, ^{2,*}riska.rts@bsi.ac.id, ³asriyani.ayy@bsi.ac.id, ⁴12230060@bsi.ac.id, ⁵12230102@bsi.ac.id

Email Penulis Korespondensi: riska.rts@bsi.ac.id

Submitted: 07/10/2024; Accepted: 31/10/2024; Published: 31/10/2024

Abstrak—Media sosial telah menjadi platform yang sangat populer dan dimanfaatkan oleh jutaan individu di seluruh dunia untuk berinteraksi serta berbagi informasi. Selain fungsinya sebagai alat komunikasi media sosial juga memainkan peran penting dalam mengekspresikan emosi dan perasaan para pengguna media sosial. Penggunaan media sosial yang meluas secara signifikan berdampak pada emosi pengguna. Emosi negatif, khususnya, sering kali muncul, yang secara drastis dapat mempengaruhi kesehatan mental. Penelitian ini bertujuan untuk membandingkan beberapa model klasifikasi untuk menemukan model terbaik dalam deteksi keseimbangan emosi pengguna media sosial. Model klasifikasi yang digunakan pada penelitian ini adalah algoritma K-Nearest Neighbor, Random Forest, Support Vector Machine, Decision Tree, dan AdaBoost untuk mengidentifikasi model klasifikasi terbaik yang mampu mendeteksi keseimbangan emosi pengguna media sosial. Beberapa model klasifikasi diterapkan dan dibandingkan dengan tujuan untuk mengevaluasi kinerja model. Penelitian ini menggunakan K-Fold Cross Validation untuk menguji model klasifikasi dengan membandingkan nilai k yang berbeda. Hasil penelitian menunjukkan bahwa algoritma Random Forest memiliki akurasi paling tinggi sebesar 99,90% pada nilai K-Fold cross validation bernilai 10 dan menghasilkan nilai Area Under Curve (AUC) sebesar 100%. Dengan demikian, penelitian ini berhasil menemukan model yang dapat diandalkan untuk mendeteksi emosi pengguna media sosial secara akurat, yang diharapkan dapat berkontribusi dalam pengembangan sistem pemantauan kesejahteraan mental di platform media sosial.

Kata Kunci: Klasifikasi; Validasi Silang; Random Forest; Multi Label; Random Forest

Abstract—Social media has grown in popularity, with millions of people using it to engage with and share information worldwide. Social media, in addition to serving as a communication tool, are crucial for expressing the emotions and feelings of users. The widespread use of social media has had a significant impact on people's emotions. In particular, negative emotions are frequently experienced and can have a significant impact on mental health. This study aimed to analyze multiple classification models to discover the optimal model for detecting emotional balance among social media users. The classification models utilized in this study include the K-Nearest Neighbor, Random Forest, Support Vector Machine, Decision Tree, and AdaBoost to identify the best classification model capable of detecting the emotional balance of social media users. Several classification models are applied and compared with the aim of evaluating model performance. This research project employed K-fold cross-validation to evaluate the categorization model by comparing various k values. The Random Forest algorithm achieved the greatest accuracy of 99.90% at a K-Fold cross validation value of 10 and an Area Under the Curve (AUC) value of 100%. Thus, this study successfully found a reliable model for accurately detecting emotions of social media users, which is expected to contribute to the development of mental well-being monitoring systems on social media platforms.

Keywords: Classification; Cross Validation; Random Forest; Multi Label; Random Forest

1. PENDAHULUAN

Media sosial secara signifikan mempengaruhi emosi pengguna, membentuk pengalaman individu dan dinamika pasar yang lebih luas. Penelitian menunjukkan bahwa emosi seperti optimisme, ketakutan, dan kegembiraan, yang berasal dari interaksi media sosial, dapat memprediksi pengembalian pasar komoditas, menyerosi arus bawah emosional dalam pengambilan keputusan keuangan.

Media sosial berfungsi sebagai platform untuk ekspresi emosional dan interaksi, mempengaruhi identitas pribadi dan penalaran emosional di antara pengguna. Selain itu, pengembangan profil emosional melalui aktivitas media sosial memungkinkan pengalaman pengguna yang disesuaikan, meningkatkan keterlibatan emosional [1], [2], [3].

Media sosial dapat menumbuhkan koneksi emosional yang positif, seperti rasa kebersamaan dan dukungan sosial, yang dapat meningkatkan kesejahteraan psikologis. Namun, di sisi lain, ia juga dapat menyebabkan keadaan emosional negatif, seperti kecemasan atau depresi, terutama ketika pengguna terlibat dalam perbandingan sosial atau mengalami cyberbullying [4], [5]. Pengaruh negatif ini sering kali diperburuk oleh dinamika sosial yang ada di platform, yang dapat memicu rasa tidak aman atau kurangnya kontrol atas citra diri. Dualitas ini menggarisbawahi hubungan kompleks antara media sosial dan emosi pengguna, menunjukkan bahwa meskipun media sosial dapat memperkaya pengalaman emosional, media sosial juga dapat membawa tantangan dalam hal keseimbangan mental dan emosional.

Keseimbangan emosi merupakan kondisi dimana seseorang mampu mengelola dan menyeimbangkan berbagai emosi yang muncul dalam kehidupan sehari-hari, yang sangat penting untuk kesehatan mental. Dalam konteks media sosial, intensitas penggunaan platform ini dapat mempengaruhi kestabilan emosi individu.



Penelitian menunjukkan bahwa semakin lama seseorang menggunakan media sosial, semakin rendah tingkat kestabilan emosinya.

Penggunaan media sosial yang berlebihan dapat menyebabkan perasaan cemas, depresi, atau stres, karena paparan terus-menerus terhadap informasi dan interaksi online yang sering kali tidak sehat. [6]. Selain itu, paparan terhadap standar kecantikan, gaya hidup ideal, dan pencapaian orang lain yang sering kali dipamerkan di media sosial dapat meningkatkan perasaan kecemasan dan ketidakpuasan diri. Penggunaan media sosial yang berlarut-larut juga dapat memperburuk kondisi emosional seseorang karena dapat menyebabkan gangguan tidur yang berpengaruh terhadap keseimbangan emosi.

Penggunaan media sosial yang berlebihan dapat menyebabkan kesepian, kecemasan, dan depresi, yang berpotensi mengganggu keseimbangan emosional individu [7], [8], [9]. Fenomena “Fear of Missing Out” (FOMO), yaitu perasaan takut tertinggal atau tidak terlibat dalam aktivitas yang dilakukan orang lain, memperburuk masalah ini, dengan menyebabkan kecanduan media sosial dan meningkatkan tekanan emosional [8].

Media sosial juga menciptakan lingkungan yang memungkinkan perbandingan sosial, yang seringkali memicu kecemburuan, ketidakpuasan dengan diri sendiri, dan penurunan harga diri [7]. Hal ini terutama terjadi karena individu cenderung hanya menunjukkan sisi positif kehidupan yang dimiliki di platform sosial, sementara sisi negatif individu sering disembunyikan. Oleh karena itu, pengguna dapat merasa kehidupan orang lain lebih bahagia atau lebih sukses, yang memengaruhi persepsi terhadap kehidupan diri sendiri.

Mendeteksi dan memonitor keadaan emosional pengguna media sosial menjadi sangat penting untuk meningkatkan kesadaran akan kesehatan mental [10]. Di sisi lain, jika digunakan dengan bijak, media sosial juga dapat memberikan dampak positif. Platform ini memungkinkan koneksi sosial, memberi kesempatan bagi orang-orang untuk berbagi pengalaman, dukungan emosional, dan mempererat hubungan yang sebelumnya terjalin jauh [11].

Bahkan, selama pandemi COVID-19, sejumlah penelitian menunjukkan bahwa penggunaan media sosial yang moderat dapat memberikan efek menguntungkan bagi kesehatan mental, membantu mengurangi rasa kesepian dan meningkatkan kebahagiaan di kalangan remaja [9]. Keseimbangan yang tepat dalam menggunakan media sosial sangat krusial untuk memaksimalkan manfaatnya tanpa menimbulkan dampak negatif pada kesehatan mental.

Berbagai penelitian yang telah dilakukan untuk deteksi keseimbangan emosi pengguna media sosial diantaranya penelitian menggunakan data teks dari ulasan pengguna media sosial [12], [13], [14] untuk mengklasifikasikan kalimat teks ke dalam kategori emosi. Beberapa penelitian lain terkait dengan klasifikasi emosi, umumnya menggunakan dataset yang berasal dari ulasan atau komentar yang ditemukan di platform media sosial Twitter [15], [16], [17].

Meskipun telah ada sejumlah penelitian yang mendalami hubungan antara penggunaan media sosial dan kesejahteraan emosional, namun masih terdapat berbagai aspek terkait yang belum terungkap secara mendalam dan banyak dari penelitian ini masih terbatas pada pengklasifikasian emosi ke dalam kategori yang umum seperti bahagia atau marah tanpa memperhitungkan kompleksitas emosi manusia. Salah satu tantangan signifikan dalam penelitian ini adalah bagaimana memanfaatkan data besar yang dihasilkan dari aktivitas pengguna di media sosial secara lebih efektif, terutama dengan menggunakan teknik machine learning untuk menghasilkan klasifikasi yang lebih tepat.

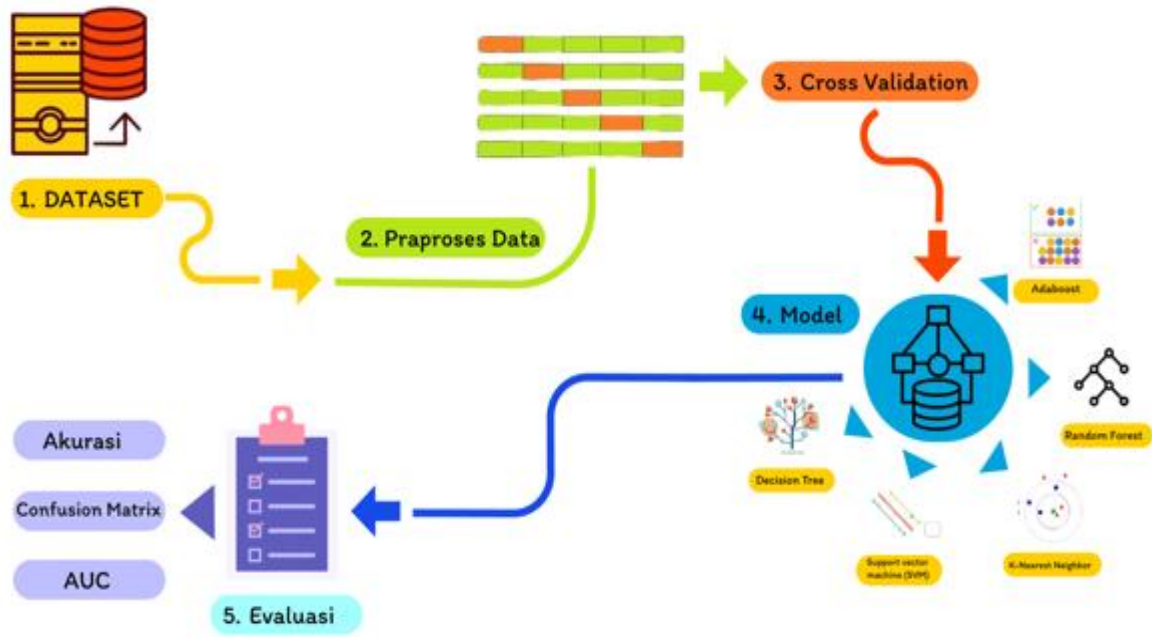
Oleh sebab itu, penelitian ini perlu dilakukan agar dapat memberikan pemahaman yang lebih mendalam tentang dinamika emosi dalam konteks media sosial dengan memanfaatkan teknik machine learning. Gap penelitian ini terletak pada kurangnya pengembangan model klasifikasi emosi yang lebih efisien dan mampu menangani kompleksitas emosi secara lebih mendalam. Penelitian ini bertujuan untuk mengembangkan model klasifikasi emosi yang lebih efisien, sehingga dapat mengolah data dalam jumlah besar secara lebih efektif. Melalui analisis data yang lebih komprehensif, penelitian ini diharapkan dapat mengidentifikasi emosi dominan yang muncul dalam interaksi pengguna, serta memahami faktor-faktor yang mempengaruhi kesejahteraan emosional pengguna.

2. METODOLOGI PENELITIAN

Penelitian ini dilakukan dengan lima tahapan berurutan, yaitu pengumpulan dataset penggunaan media sosial dan kesejahteraan emosional, praproses data, pembagian data untuk pelatihan dan pengujian menggunakan cross validation, pelatihan model, dan analisis hasil evaluasi. Tahapan penelitian ini dapat dilihat pada gambar 1.

2.1 Dataset

Penelitian ini menggunakan dataset penggunaan media sosial dan kesejahteraan emosional yang diambil dari Kaggle [18]. Dataset ini berisi penggunaan media sosial dan keadaan emosi dominan pengguna berdasarkan aktivitas yang dilakukan di media sosial. Dataset ini terdiri dari 1.000 data. Dataset penggunaan media sosial dan kesejahteraan emosional terdiri dari 10 atribut dengan 9 atribut prediksi dan 1 atribut target yang dapat dilihat pada tabel 1. Target yang digunakan adalah keadaan emosi dominan pengguna media sosial yang mempunyai enam label, yaitu kebahagiaan, kesedihan, marah, kecemasan, kebosanan, dan netral.



Gambar 1. Tahapan Penelitian

2.2 Praproses Data

Praproses data merupakan langkah penting yang harus dilakukan agar data yang digunakan tidak ada noise [19], [20]. Tahap awal dilakukan dengan menangani data yang hilang dan menghapus atribut yang berisi informasi pribadi. Atribut yang tidak digunakan, yaitu User ID.

2.3 Cross Validation

Cross validation memiliki peran penting dalam proses pemilihan dan evaluasi model pembelajaran mesin dengan menyediakan pendekatan sistematis untuk menilai kinerja model. K-fold Cross Validation memungkinkan pemilihan model yang lebih kuat dengan merata-ratakan kinerja di beberapa fold, yang dapat sangat bermanfaat dalam skenario data yang kompleks, seperti yang melibatkan banyak struktur [21]. Cross Validation membantu memperkirakan seberapa baik kinerja model pada data yang tidak terlihat dengan membagi dataset ke dalam data latih dan data uji, sehingga mengurangi risiko overfitting [22]. K-fold cross validation membantu dalam menyetel parameter model dan menilai kinerja, sehingga menyediakan kerangka evaluasi yang kuat untuk tugas klasifikasi multi-label [23]. Pada penelitian ini, nilai K-Fold Cross Validation yang digunakan adalah 5, 10, dan 20. Pada nilai k=10, data dibagi menjadi 10-fold dengan ukuran sama, dimana 9-fold akan digunakan sebagai data latih dan 1-fold digunakan sebagai data uji seperti pada gambar 2.

Tabel 1. Deskripsi Atribut Dataset

No	Atribut	Tipe Data	Keterangan
1	User ID	Categorical	Pengidentifikasi unik dari pengguna
2	Age	Numeric	Usia pengguna media sosial
3	Gender	Categorical	Jenis kelamin pengguna
4	Platform	Categorical	Platform media sosial yang digunakan
5	Daily_Usage_Time (minutes)	Numeric	Waktu harian yang dihabiskan di platform dalam hitungan menit
6	Posts_Per_Day	Numeric	Jumlah postingan yang dibuat per hari
7	Likes_Received_Per_Day	Numeric	Jumlah suka yang diterima per hari
8	Comments_Received_Per_Day	Numeric	Jumlah komentar yang diterima per hari
9	Messages_Sent_Per_Day	Numeric	Jumlah pesan yang dikirim per hari
10	Dominant_Emotion	Categorical	Keadaan emosi dominan pengguna sepanjang hari

2.4 Model

Klasifikasi kesejahteraan emosional melalui penggunaan media sosial dilakukan dengan algoritma pembelajaran mesin seperti K-Nearest Neighbor (KNN), Decision Tree, Support Vector Machine (SVM), Random Forest (RF), dan AdaBoost.



Gambar 2. 10 Fold Cross Validation

- a. K-Nearest Neighbor (KNN)
 Algoritma ini menentukan kelas dari sampel data baru berdasarkan mayoritas kelas dari data tetangga terdekatnya [24], [25].
- b. Decision Tree
 Metode ini merepresentasikan keputusan dan konsekuensinya dalam bentuk struktur pohon, dimana setiap cabang mewakili pilihan atau aturan yang digunakan untuk membuat keputusan [26], [27], [28]
- c. Support Vector Machine (SVM)
 Algoritma SVM bertujuan menemukan hyperplane yang paling optimal, yaitu garis pemisah terbaik, untuk membagi data dari kelas-kelas yang berbeda [29].
- d. Random Forest (RF)
 Algoritma ini memanfaatkan banyak pohon keputusan (decision tree) secara bersamaan untuk memperbaiki akurasi prediksi, dengan cara mengkombinasikan hasil dari tiap pohon [29].
- e. AdaBoost
 Sebuah metode boosting yang menggabungkan sejumlah model prediksi sederhana (weak learners) untuk memperbaiki keseluruhan performa prediksi [30].

2.5 Evaluasi

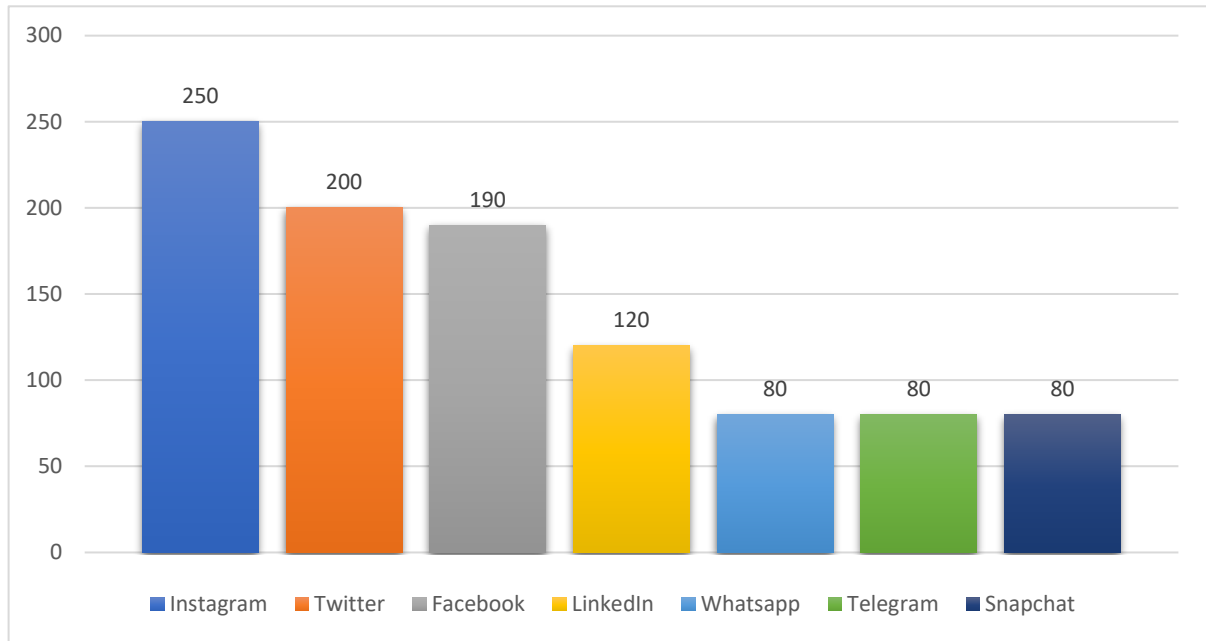
Evaluasi model dilakukan berdasarkan akurasi, confusion matrix, dan AUC. Akurasi adalah metrik sederhana yang menghitung persentase prediksi benar dari keseluruhan prediksi. Untuk memahami hasil klasifikasi lebih dalam, confusion matrix digunakan sebagai alat evaluasi yang memberikan gambaran rinci tentang prediksi model. Confusion matrix menunjukkan distribusi prediksi dalam kategori: True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). AUC adalah skor probabilitas yang mengukur kemampuan model untuk membedakan antara kelas positif dan negatif di berbagai ambang batas prediksi. Nilai AUC berkisar dari 0 hingga 1, di mana semakin mendekati 1, semakin baik model dalam membedakan kedua kelas. AUC sangat bermanfaat pada kasus dengan kelas tidak seimbang karena mengukur sensitivitas model terhadap kelas positif tanpa bias terhadap distribusi kelas.

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan 1000 data yang berasal dari dataset penggunaan media sosial dan kesejahteraan emosional yang diperoleh dari Kaggle [18]. Data ini menyediakan informasi tentang bagaimana penggunaan media sosial dapat berdampak pada kesejahteraan emosional individu. Variabel dalam dataset ini mencakup berbagai aspek penggunaan media sosial, seperti frekuensi penggunaan, platform yang paling sering digunakan, serta durasi waktu yang dihabiskan. Selain itu, data ini juga mencakup aspek kesejahteraan emosional, seperti kesedihan, kecemasan, dan kebahagiaan yang dirasakan oleh pengguna media sosial.

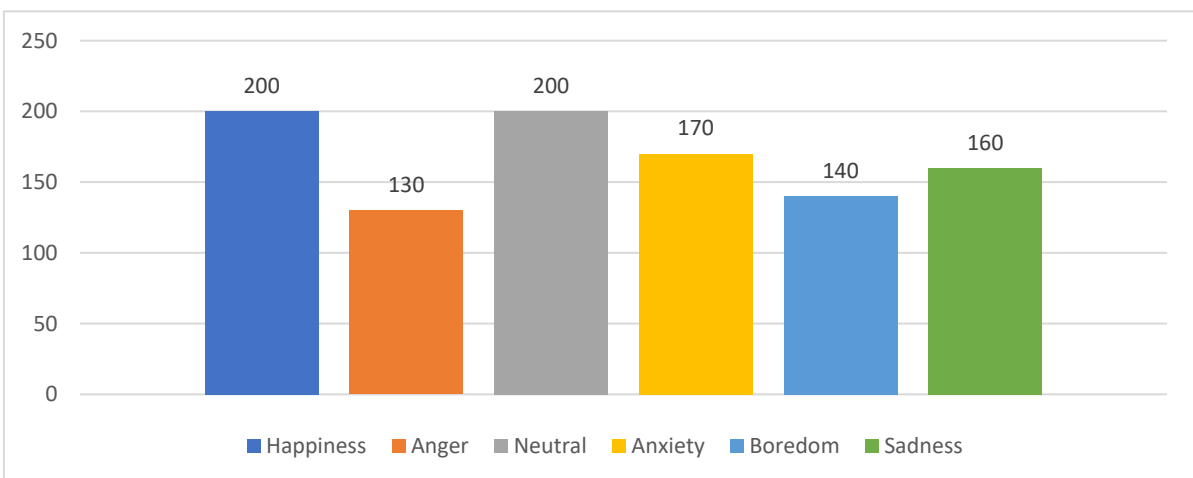
Platform media sosial yang digunakan adalah instagram, twitter, facebook, linkedin, whatsapp, telegram, dan snapchat. Berdasarkan gambar 2 menunjukkan perbandingan penggunaan beberapa platform media sosial.

Instagram menjadi platform yang paling banyak digunakan, dengan angka tertinggi yaitu 250, menunjukkan popularitasnya yang besar dibandingkan platform lain. Di posisi kedua, Twitter mencatatkan angka 200, diikuti oleh Facebook yang sedikit lebih rendah pada 190. LinkedIn, sebagai platform profesional, berada di tengah dengan angka 120. Sementara itu, tiga platform lainnya, WhatsApp, Telegram, dan Snapchat, memiliki jumlah yang sama, masing-masing sebesar 80, yang menunjukkan bahwa meskipun platform cukup populer, penggunaannya masih berada di bawah platform utama lainnya seperti Instagram, Twitter, dan Facebook. Data ini mengindikasikan bahwa media sosial yang lebih visual atau berbasis komunikasi cenderung lebih diminati oleh pengguna.



Gambar 2. Platform Media Sosial

Keadaan emosi dominan pengguna media sosial ditunjukkan pada gambar 3 yang menampilkan berbagai emosi dan frekuensi kemunculannya. Happiness dan Neutral memiliki frekuensi tertinggi, masing-masing dengan angka 200, menunjukkan bahwa kedua emosi ini lebih mendominasi. Anxiety (kecemasan) berada di posisi berikutnya dengan nilai 170, menunjukkan bahwa kecemasan juga merupakan perasaan yang cukup umum. Di sisi lain, Sadness (kesedihan) mencatat angka 160, sedikit di bawah kecemasan, yang mengindikasikan bahwa perasaan sedih cukup sering terjadi. Boredom (kebosanan) memiliki angka 140, lebih rendah dari kecemasan dan kesedihan, tetapi masih signifikan. Sementara itu, Anger (kemarahan) mencatat nilai terendah dengan angka 130, menjadikannya emosi yang paling jarang muncul di antara kategori lainnya. Gambar 3 menunjukkan keseimbangan antara emosi positif, netral, dan negatif dalam frekuensi kemunculannya.



Gambar 3. Emosi Dominan Pengguna Media Sosial

Data yang telah diperoleh dibagi menjadi data pelatihan dan data pengujian menggunakan K-Fold Cross Validation dengan nilai k yang berbeda, yaitu k=5, k=10, dan k=20. Data pelatihan dan pengujian dimasukkan ke

dalam algoritma K-Nearest Neighbor (KNN), Decision Tree, Support Vector Machine (SVM), Random Forest (RF), serta AdaBoost.

Tabel 2 menunjukkan evaluasi lima model machine learning yang berbeda, yaitu KNN, Decision Tree, SVM, Random Forest, dan AdaBoost, berdasarkan metrik akurasi dan AUC dengan variasi K-Fold (5, 10, 20). Untuk model KNN, akurasi stabil di angka 0,976 pada semua variasi K-Fold, dengan AUC yang sedikit meningkat dari 0,987 menjadi 0,988. Decision Tree memiliki akurasi yang bervariasi, mulai dari 0,959 pada 5 K-Fold hingga 0,967 pada 20 K-Fold, dengan AUC yang sedikit meningkat dari 0,981 menjadi 0,983. Model SVM memiliki akurasi yang lebih rendah dibandingkan model lainnya, berkisar antara 0,912 hingga 0,917, namun AUC-nya sangat tinggi, mencapai 0,996 pada K-Fold 10.

Tabel 2. Hasil Evaluasi Model

Model	K-Fold	Akurasi	AUC
KNN	5	0,976	0,987
	10	0,976	0,988
	20	0,976	0,988
Decision Tree	5	0,959	0,981
	10	0,966	0,982
	20	0,967	0,983
SVM	5	0,912	0,995
	10	0,917	0,996
	20	0,913	0,996
Random Forest	5	0,987	1,000
	10	0,990	1,000
	20	0,988	1,000
AdaBoost	5	0,969	0,981
	10	0,969	0,981
	20	0,966	0,9761

Random Forest terbukti sebagai model dengan performa terbaik, dengan akurasi tertinggi mencapai 0,990 pada 10 K-Fold dan AUC sempurna 1,000 di semua variasi K-Fold. Model AdaBoost memiliki performa yang relatif konsisten dengan akurasi sekitar 0,969, meskipun AUC-nya sedikit menurun dari 0,981 menjadi 0,979 pada 20 K-Fold. Sementara itu, model SVM juga menunjukkan hasil yang baik, meskipun tidak sebaik Random Forest dalam hal akurasi. Dari hasil ini, dapat disimpulkan bahwa Random Forest adalah model yang paling akurat untuk data penggunaan media sosial dan kesejahteraan emosional, sementara model lain seperti SVM dan AdaBoost juga menunjukkan hasil yang baik, terutama dalam AUC.

Tabel 3 merupakan confusion matrix dari model Random Forest dengan K-Fold Cross Validation (nilai K = 10) yang digunakan untuk mengklasifikasikan berbagai emosi. Performa model dalam mengklasifikasikan emosi seperti anger (marah), anxiety (kecemasan), boredom (bosan), happiness (kebahagiaan), neutral (netral), dan sadness (kesedihan) dievaluasi dengan membandingkan prediksi model terhadap data aktual. Model mampu memprediksi 129 dari 130 data anger secara benar, dengan hanya 1 data yang salah diklasifikasikan sebagai sadness (kesedihan). Ini menunjukkan bahwa model memiliki performa yang sangat baik dalam mengklasifikasikan emosi anger. Sebanyak 165 data anxiety diprediksi dengan benar, yang menunjukkan bahwa model sangat akurat dalam mengklasifikasikan emosi anxiety. Kelas boredom, 138 data dari 140 data diprediksi dengan benar oleh model, sementara 1 data diprediksi sebagai sadness dan 1 data diprediksi sebagai anxiety. Model mampu mengklasifikasikan 198 dari 200 data happiness dengan benar, tetapi 2 data salah diklasifikasikan sebagai neutral. Ini menunjukkan bahwa ada sedikit kebingungan antara emosi happiness dan neutral, namun performa model tetap sangat baik. Model berhasil memprediksi semua data neutral dengan benar sebanyak 200 data, yang juga menunjukkan tingkat akurasi yang sangat tinggi dalam mendeteksi emosi neutral. Model secara sempurna mengklasifikasikan 160 data sadness tanpa kesalahan, yang menunjukkan kemampuan model yang sangat akurat dalam mendeteksi emosi ini. Secara keseluruhan, model Random Forest menunjukkan performa yang sangat baik dalam mengklasifikasikan emosi dengan tingkat kesalahan yang sangat rendah. Ada sedikit kebingungan antara kategori happiness dan neutral, yang mungkin disebabkan oleh kesamaan ekspresi atau ciri emosional dalam data tersebut, namun secara keseluruhan model ini memiliki kemampuan klasifikasi yang sangat baik berdasarkan confusion matrix ini.

Tabel 3. Confusion Matrix Random Forest dengan nilai K-Fold = 10

A c t u a	Predicted					
	Anger	Anxiety	Boredom	Happiness	Neutral	Sadness
Anger	129	0	0	0	0	1
Anxiety	0	165	2	2	0	1
Boredom	0	1	138	0	0	1



a	Happiness	0	0	0	198	2	0
l	Neutral	0	0	0	0	200	0
	Sadness	0	0	0	0	0	160

4. KESIMPULAN

Pada penelitian ini, berbagai algoritma pembelajaran mesin yang digunakan dalam mendeteksi keseimbangan emosi pengguna media sosial berdasarkan dataset penggunaan media sosial dan kesejahteraan emosional. Algoritma yang digunakan dalam penelitian ini adalah K-Nearest Neighbor (KNN), Decision Tree, Support Vector Machine (SVM), Random Forest (RF), dan AdaBoost. Evaluasi model dilakukan menggunakan Teknik K-Fold Cross Validation dengan variasi nilai k untuk mengukur kinerja masing-masing model. Hasil evaluasi menunjukkan bahwa algoritma Random Forest terbukti memberikan performa terbaik dengan akurasi tertinggi sebesar 99.90% dan nilai AUC sebesar 100% sehingga menjadikan random forest sebagai model yang paling efektif dalam klasifikasi multi label emosi. Selain random forest, AdaBoost dan SVM juga menunjukkan hasil yang baik dengan nilai AUC tinggi. Meskipun keduanya memiliki akurasi sedikit lebih rendah dibandingkan random forest. Algoritma KNN memberikan performa yang stabil dan konsisten namun tidak mencapai tingkat akurasi yang tinggi seperti yang diperoleh oleh Random Forest dan AdaBoost. Sementara itu, Decision Tree menunjukkan variasi hasil yang lebih besar tergantung pada nilai k pada K-Fold, yang dapat mempengaruhi kestabilan model. Secara keseluruhan, penelitian ini dapat disimpulkan bahwa random forest dan AdaBoost, memiliki keunggulan dalam menangani data yang kompleks dan tidak seimbang, serta mampu memberikan hasil prediksi yang lebih akurat dibandingkan algoritma KNN, decision tree, dan SVM terutama dalam konteks deteksi emosi pengguna media sosial.

UCAPAN TERIMAKASIH

Penulis menyampaikan apresiasi yang mendalam terhadap Universitas Bina Sarana Informatika atas dukungan pendanaan yang diberikan pada penelitian ini dengan skema Penelitian Dana Yayasan (PDY). Semoga hasil penelitian ini dapat memberikan kontribusi yang berarti bagi perkembangan ilmu pengetahuan dan bermanfaat bagi masyarakat luas.

REFERENCES

- [1] M. Schreiner, T. Fischer, and R. Riedl, "Impact of Content Characteristics and Emotion on Behavioral Engagement in Social Media: Literature Review and Research Agenda," *Electronic Commerce Research*, vol. 21, no. 2, pp. 329–345, Jun. 2021, doi: 10.1007/s10660-019-09353-8.
- [2] N. Hicham, S. Karim, and N. Habbat, "Customer sentiment analysis for Arabic social media using a novel ensemble machine learning approach," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 4, pp. 4504–4515, Aug. 2023, doi: 10.11591/ijece.v13i4.pp4504-4515.
- [3] R. Dolan, J. Conduit, C. Frethey-Bentham, J. Fahy, and S. Goodman, "Social Media Engagement Behavior," *Eur J Mark*, vol. 53, no. 10, pp. 2213–2243, Oct. 2019, doi: 10.1108/EJM-03-2017-0182.
- [4] J. A. Naslund, A. Bondre, J. Torous, and K. A. Aschbrenner, "Social Media and Mental Health: Benefits, Risks, and Opportunities for Research and Practice," *J Technol Behav Sci*, vol. 5, no. 3, pp. 245–257, Sep. 2020, doi: 10.1007/s41347-020-00134-x.
- [5] M. A. Moreno and A. F. Jolliff, "Depression and Anxiety in the Context of Digital Media," in *Handbook of Adolescent Digital Media Use and Mental Health*, Cambridge University Press, 2022, pp. 217–241. doi: 10.1017/9781108976237.013.
- [6] S. Salsabillah, N. R. Siregar, and Y. A. Pambudhi, "Intensitas Penggunaan Media Sosial dengan Kestabilan Emosi pada Mahasiswa Kedokteran," *Jurnal Sublimapsi*, vol. 4, no. 2, p. 247, May 2023, doi: 10.36709/sublimapsi.v4i2.29321.
- [7] K. Rani, F. Fatima, and A. Kumar, "Influence of Social Media on Educational and Mental Wellbeing of Young Minds: A Critical Analysis," *International Journal For Multidisciplinary Research*, vol. 6, no. 4, Jul. 2024, doi: 10.36948/ijfmr.2024.v06i04.24666.
- [8] N. Pacocha and O. Gugała, "The Influence of Social Media on Mental Wellbeing - A Review of Literature," *International Journal of Innovative Technologies in Social Science*, no. 2(42), May 2024, doi: 10.31435/rsglobal_ijitss/30062024/8151.
- [9] A. Mulder, K. Kingsley, D. Camacho, and J. Lasprilla, "Adolescent Emotional Well-Being and Social Media Addiction: The COVID-19 Pandemic's Influence on Mental Health," *Med Res Arch*, vol. 12, no. 5, 2024, doi: 10.18103/mra.v12i5.5289.
- [10] F. A. Acheampong, C. Wenyu, and H. Nunoo-Mensah, "Text-Based Emotion Detection: Advances, Challenges, and Opportunities," *Engineering Reports*, vol. 2, no. 7, Jul. 2020, doi: 10.1002/eng2.12189.
- [11] M. L. BĂRHALESCU and E. COSTESCU, "The Role of Social Media in User Health," *Journal of Marine Technology and Environment*, vol. 2, no. 2, pp. 7–11, Oct. 2023, doi: 10.53464/JMTE.02.2023.01.
- [12] A. M. Abubakar, D. Gupta, and S. Palaniswamy, "Explainable Emotion Recognition from Tweets using Deep Learning and Word Embedding Models," in *2022 IEEE 19th India Council International Conference (INDICON)*, IEEE, Nov. 2022, pp. 1–6. doi: 10.1109/INDICON56171.2022.10039878.
- [13] P. Kumar and B. Raman, "A BERT Based Dual-Channel Explainable Text Emotion Recognition System," *Neural Networks*, vol. 150, pp. 392–407, Jun. 2022, doi: 10.1016/j.neunet.2022.03.017.



- [14] Y. Astari, A. Afiyati, and S. W. Rozaqi, "Analisis Sentimen Multi-Class pada Sosial Media menggunakan metode Long Short-Term Memory (LSTM)," *Jurnal Linguistik Komputasional*, vol. 4, no. 1, pp. 8–12, 2021, doi: 10.26418/jlk.v4i1.43.
- [15] S. Sudiarto, "Analisis Kinerja Algoritma Machine Learning Untuk Klasifikasi Emosi," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 2, Sep. 2022, doi: 10.47065/bits.v4i2.2261.
- [16] I. G. Harsemadi, I. K. Dharmendra, and I. M. P. P. Wijaya, "Klasifikasi Emosi Pada Tweet Berbahasa Indonesia Menggunakan Teknik Sampling ENN," *Jurnal Teknologi Informasi dan Komputer*, vol. 9, no. 5, Oct. 2023, doi: 10.36002/jutik.v9i5.2646.
- [17] A. Sujjada and Anggun Fergina, "Implementasi Metode Vector Space Model untuk Deteksi Emosi Menggunakan Data Teks Twitter," *Jurnal RESTIKOM : Riset Teknik Informatika dan Komputer*, vol. 3, no. 3, pp. 116–129, Jun. 2022, doi: 10.52005/restikom.v3i3.89.
- [18] E. Bulut, "Social Media Usage and Emotional Well-being," Kaggle. Access Date Sept 2024 [Online]. Available: <https://www.kaggle.com/datasets/emirhanai/social-media-usage-and-emotional-well-being>
- [19] V. Kumar, S. Kumar, and S. Sarangi, "Effect of Sampling Rate on Parametric and Non-parametric Data Preprocessing for Gearbox Fault Diagnosis," *Journal of Vibration Engineering & Technologies*, vol. 12, no. 2, pp. 1195–1202, Feb. 2024, doi: 10.1007/s42417-023-00901-z.
- [20] T. I. Fajri et al., *Data Mining*. Payakumbuh: PT. Serasi Media Teknologi, 2024.
- [21] D. Brown, "Model Selection Through Cross-Validation for Supervised Learning Tasks with Manifold Data," *The Journal of Purdue Undergraduate Research*, vol. 13, no. 1, Jan. 2024, doi: 10.7771/2158-4052.1585.
- [22] V. Q. Noor, R. Herfiansyah, A. S. Ramadhan, and A. Amali, "Prediksi Diabetes Menggunakan Algoritma Naive Bayes Menggunakan Rapidminer," *Pelita Teknologi*, vol. 19, no. 1, pp. 16–21, Mar. 2024, doi: 10.37366/pelitatekno.v19i1.4329.
- [23] A. Muhaimin, W. Wibowo, and P. A. Riyantoko, "Multi-label Classification Using Vector Generalized Additive Model via Cross-Validation," *Journal of Information and Communication Technology*, vol. 22, 2023, doi: 10.32890/jict2023.22.4.5.
- [24] I. Kurniawan and P. B. Santoso, "Design of K-Nearest Neighbor Algorithm For Classification of Credit Loan Eligibility At Senarak Dana Purwakarta Cooperative," *Internet of Things and Artificial Intelligence Journal*, vol. 4, no. 2, pp. 354–370, Jun. 2024, doi: 10.31763/iota.v4i2.742.
- [25] K. Iyer, A. Shukla, K. Sharma, and M. Varghese, "Speech Emotion Recognition using Gaussian Mixture Model (GMM) and K-Nearest Neighbors (KNN)," in *Advancements in Communication and Systems*, Soft Computing Research Society, 2024, pp. 443–455. doi: 10.56155/978-81-955020-7-3-39.
- [26] P. P. Putra, M. K. Anam, S. Defit, and A. Yuniarta, "Enhancing the Decision Tree Algorithm to Improve Performance Across Various Datasets," *INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi*, vol. 8, no. 2, pp. 200–212, Aug. 2024, doi: 10.29407/intensif.v8i2.22280.
- [27] N. Celiker and C. O. Guzeller, "Predicting Organizational Citizenship Behaviour in Hospitality Businesses with Decision Tree Method," *International Journal of Hospitality & Tourism Administration*, vol. 25, no. 2, pp. 436–474, Mar. 2024, doi: 10.1080/15256480.2022.2120942.
- [28] I. D. Mienye and N. Jere, "A Survey of Decision Trees: Concepts, Algorithms, and Applications," *IEEE Access*, vol. 12, pp. 86716–86727, 2024, doi: 10.1109/ACCESS.2024.3416838.
- [29] R. F. Putra et al., *Algoritma Pembelajaran Mesin: Dasar, Teknik, dan Aplikasi*. Jambi: PT. Sonpedia Publishing Indonesia, 2024.
- [30] S. Madanian et al., "Speech emotion recognition using machine learning - A systematic review," *Intelligent Systems with Applications*, vol. 20, p. 200266, Nov. 2023, doi: 10.1016/j.iswa.2023.200266.