

SKRIPSI

**PERBANDINGAN METODE *K-NEAREST NEIGHBOR* (KNN)
DAN *RANDOM FOREST* UNTUK KLASIFIKASI SENTIMEN
KOMENTAR DI TIKTOK**

(Studi Kasus: Performa ONIC Esports pada turnamen *M7 Mobile Legends: Bang Bang*)



ANGGIT PRATIKNYO WIDYAGIRI

NIM. 22106010075

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA**

2026

**PERBANDINGAN METODE *K-NEAREST NEIGHBOR* (KNN)
DAN *RANDOM FOREST* UNTUK KLASIFIKASI SENTIMEN
KOMENTAR DI TIKTOK**

(Studi Kasus: Performa ONIC Esports pada turnamen *M7 Mobile Legends: Bang Bang*)

Skripsi

Untuk memenuhi sebagian persyaratan

Mencapai derajat Sarjana S-1

Program Studi Matematika



Diajukan oleh

ANGGIT PRATIKNYO WIDYAGIRI

22106010075

Kepada

PROGRAM STUDI MATEMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA

YOGYAKARTA

2026



PENGESAHAN TUGAS AKHIR

Nomor : B-2044/Un.02/DST/PP.00.9/08/2026

Tugas Akhir dengan judul : Perbandingan Metode K-Nearest Neighbor (KNN) dan Random Forest untuk Klasifikasi Sentimen Komentar di TikTok (Studi Kasus: Performa ONIC Esports pada turnamen M7 Mobile Legends: Bang Bang)

yang dipersiapkan dan disusun oleh:

Nama : ANGGIT PRATIKNYO WIDYAGIRI
Nomor Induk Mahasiswa : 22106010075
Telah diujikan pada : Selasa, 11 Agustus 2026
Nilai ujian Tugas Akhir : A/B

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang

Muchammad Abrori, S.Si., M.Kom
SIGNED

Valid ID: 6a88013e76540



Penguji I

Muhamad Rashif Hilmi, S.Si., M.Sc.
SIGNED

Valid ID: 6a88178033621



Penguji II

Sri Utami Zuliana, S.Si., M.Sc., Ph.D.
SIGNED

Valid ID: 6a87f905bb15f



Yogyakarta, 11 Agustus 2026
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi
Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 6a8908ab95022

SURAT PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini:

Nama : Anggit Pratiknyo Widyagiri

NIM : 22106010075

Program Studi : Matematika

Fakultas : Sains dan Teknologi

Dengan ini menyatakan bahwa isi skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar sarjana di suatu Perguruan Tinggi dan sesungguhnya skripsi ini merupakan hasil pekerjaan penulis sendiri sepanjang pengetahuan penulis, bukan duplikasi atau saduran dari karya orang lain kecuali bagian tertentu yang penulis ambil sebagai bahan acuan. Apabila terbukti pernyataan ini tidak benar, sepenuhnya menjadi tanggung jawab penulis.

Yogyakarta, 21 Agustus 2026



Anggit Pratiknyo Widyagiri

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA



SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi / Tugas Akhir

Lamp : -

Kepada

Yth. Dekan Fakultas Sains dan Teknologi
UTN Sunan Kalijaga Yogyakarta
di Yogyakarta

Assalamu'alaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Anggit Pratiknyo Widyagiri

NIM : 22106010075

Judul Skripsi : Perbandingan Metode K-Nearest Neighbor (KNN) dan Random Forest untuk Klasifikasi Sentimen Komentar di TikTok (Studi Kasus: Performa ONIC Esports pada Turnamen M7 Mobile Legends: Bang Bang)

sudah dapat diajukan kembali kepada Program Studi Matematika Fakultas Sains dan Teknologi UTN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Matematika.

Dengan ini kami mengharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqasyahkan. Atas perhatiannya kami ucapkan terima kasih.

Wassalamu'alaikum wr. wb.

Yogyakarta, 04 Agustus 2026

Pembimbing I

Muchammad Abrori, S.Si., M.Kom
NIP. 197204231999031003

Pembimbing II

Muhamad Rashif Hilmi, S.Si., M.Sc.
NIP. 199203092020121001

MOTTO

“If you can’t do great things, do small things in a great way”

(Napoleon Hill)

PERSEMBAHAN

*Dengan penuh rasa syukur, skripsi ini
saya persembahkan kepada
kedua orang tua tercinta,
kakak dan adik tersayang,
keluarga besar,
serta teman-teman seperjuangan.
Terima kasih atas doa, kasih sayang,
dukungan, dan semangat yang diberikan
hingga akhirnya karya ini dapat diselesaikan.*



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

PRAKATA

Alhamdulillah rabbil'alamiin, segala puji bagi Allah SWT atas limpahan rahmat, taufik, dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Perbandingan Metode *K-Nearest Neighbor* (KNN) dan *Random Forest* untuk Klasifikasi Sentimen Komentar di TikTok (Studi Kasus: Performa ONIC Esports pada turnamen M7 *Mobile Legend: Bang Bang* (MLBB)”. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Matematika pada Program Studi Matematika, Fakultas Sains dan Teknologi, UIN Sunan Kalijaga Yogyakarta. Sholawat serta salam semoga senantiasa tercurah kepada Nabi Muhammad SAW, beserta keluarga, sahabat, dan seluruh umatnya hingga akhir zaman.

Penulis menyadari bahwa proses penyusunan skripsi ini tidak terlepas dari bantuan, bimbingan, arahan, serta dukungan dari berbagai pihak. Oleh karena itu, dengan penuh rasa hormat penulis menyampaikan terimakasih kepada:

1. Prof. H. Noorhaidi, S.Ag., M.A., M.Phil., Ph.D., selaku Rektor UIN Sunan Kalijaga.
2. Prof. Dr. Khurul Wardati, S.Si., M.Si., selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga.
3. Dr. Epha Diana Supandi, S.Si., M.Sc., selaku Ketua Program Studi Matematika.
4. Muchammad Abrori, S.Si., M.Kom., selaku dosen pembimbing 1 yang telah membimbing, mengarahkan, dan memotivasi penulis dalam penyusunan skripsi ini sehingga dapat terselesaikan dengan baik.
5. Muhammad Rashif Hilmi, S.Si., M.Sc., selaku dosen pembimbing 2 yang telah membimbing, mengarahkan, dan memotivasi penulis dalam penyusunan skripsi ini sehingga dapat terselesaikan dengan baik.
6. Seluruh Bapak/Ibu Dosen Program Studi Matematika yang telah membekali penulis dengan ilmu pengetahuan, pengalaman, dan

keteladanan selama masa perkuliahan. Semoga Allah SWT membalas segala kebaikan yang telah diberikan.

7. Kedua orang tua penulis, Bapak Aditya Widyagiri dan Ibu Poppy Dian Fardiah, atas doa, kasih sayang, serta dukungan lahir dan batin yang tidak pernah putus. Adik penulis Triasmo Bayudono Widyagiri, dan Kakak penulis Ariobimo Priambodo Widyagiri yang selalu memberikan semangat kepada penulis dalam menyelesaikan studi.
8. Teman terdekat penulis Derrida, serta seluruh mahasiswa Matematika angkatan 2022 yang telah menjadi teman berbagi suka dan duka, kasih sayang, dan dukungan yang diberikan sampai saat ini.
9. Serta semua pihak yang telah membantu penulis dalam penyusunan skripsi ini yang tidak dapat penulis sebutkan satu per satu.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan dan jauh dari kata sempurna. Oleh karena itu, penulis memohon maaf atas segala kekurangan dan sangat mengharapkan kritik serta saran yang membangun demi kebaikan di masa mendatang. Semoga skripsi ini dapat memberikan manfaat serta menambah wawasan, bagi penulis maupun pembaca.

Yogyakarta, 5 Agustus 2026

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
SURAT PERNYATAAN KEASLIAN	iii
SURAT PERSETUJUAN SKRIPSI	iv
MOTTO	v
PERSEMBAHAN	vi
PRAKATA	vii
DAFTAR ISI	ix
INTISARI	xiii
BAB I	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Tujuan Penelitian	5
1.4 Batasan Masalah	6
1.5 Manfaat Penelitian	6
1.6 Tinjauan Pustaka	7
1.7 Sistematika Penulisan	9
BAB II	12
LANDASAN TEORI	12
2.1 Vektor	12
2.2 Matriks	13
2.3 Jarak	14
2.4 <i>Data mining</i>	15
2.5 <i>Text Mining</i>	16
2.6 Analisis Sentimen	17
2.7 <i>Preprocessing</i>	18
2.8 Representasi Teks	20
2.9 <i>Vector Space Model</i>	20
2.10 TF-IDF	22

2.11	<i>Machine Learning</i>	23
2.12	<i>Supervised Learning</i>	24
2.13	Klasifikasi.....	25
2.14	<i>Decision Tree</i>	26
2.15	<i>k-Fold Cross Validation</i>	29
2.16	<i>Grid Search</i>	31
2.17	<i>Confusion Matrix</i>	31
BAB III		35
METODE PENELITIAN		35
3.1	Jenis Penelitian	35
3.2	Metode Pengumpulan Data	36
3.3	Variabel Penelitian	37
3.4	Metode Penelitian.....	37
3.1	Metode Analisis Data	38
3.5	Alur Penelitian.....	40
BAB IV		41
PEMBAHASAN		41
4.1	Metode <i>k-Nearest Neighbor</i> (kNN).....	41
4.1.1	Pengertian dan Konsep Dasar	41
4.1.2	Pemilihan Nilai Parameter <i>k</i>	42
4.1.3	Cara Kerja kNN.....	44
4.1.4	Pengujian Model kNN	46
4.2	Metode <i>Random Forest</i>	47
4.2.1	Pengertian dan Konsep Dasar	47
4.2.2	Parameter pada <i>Random Forest</i>	49
4.2.3	Cara Kerja <i>Random Forest</i>	50
4.2.4	Pengujian Model <i>Random Forest</i>	52
4.3	Contoh Perhitungan Manual Metode kNN dan <i>Random Forest</i>	53
4.3.1	Contoh Perhitungan manual metode kNN	53
4.3.2	Contoh perhitungan manual metode <i>Random Forest</i>	58
BAB V		65
STUDI KASUS		65

5.1	Pengumpulan Data	65
5.2	Pelabelan Data	66
5.3	<i>Data Preprocessing</i>	68
5.3.1	<i>Cleaning</i>	68
5.3.2	<i>Case Folding</i>	69
5.3.3	Tokenisasi	69
5.3.4	Normalisasi	70
5.3.5	<i>Stopword Removal</i>	71
5.3.6	<i>Stemming</i>	72
5.4	<i>Data Splitting</i>	73
5.5	Ekstraksi Fitur	74
5.6	Pembentukan dan Pengujian Model	77
5.6.1	Pembentukan Model <i>k-Nearest Neighbor</i> (kNN)	78
5.6.2	Pembentukan dan Pengujian Model kNN (<i>Euclidean</i>)	79
5.6.3	Pembentukan dan Pengujian Model kNN (<i>Manhattan</i>)	80
5.6.4	Pembentukan dan Pengujian Model <i>Random Forest</i>	82
5.7	Evaluasi Kinerja Metode kNN dan Random Forest	85
BAB VI		88
PENUTUP		88
6.1	Kesimpulan	88
6.2	Saran	89
DAFTAR PUSTAKA		90
LAMPIRAN		94
Lampiran-1 Codingan <i>scraping</i> data		94
Lampiran-2 Codingan penyaringan data berbahasa Indonesia		98
Lampiran-3 Codingan penggabungan seluruh hasil <i>scraping</i>		100
Lampiran-4 Codingan pelabelan data		102
Lampiran-5 Codingan seluruh tahapan <i>Preprocessing</i>		105
Lampiran-6 Codingan pembagian data latih dan data uji		112
Lampiran-7 Codingan ekstraksi fitur dengan TF-IDF		113
Lampiran-7.1 Codingan penerapan TF-IDF pada 5 komentar		113
Lampiran-8 Codingan pembentukan dan pengujian model kNN (<i>Euclidean</i>)		115

Lampiran-9 Codingan pembentukan dan pengujian model kNN (<i>Manhattan</i>)	117
Lampiran-10 Codingan pembentukan dan pengujian model <i>Random Forest</i> .	118
CURRICULUM VITAE	121



INTISARI

Perbandingan Metode *K-Nearest Neighbor* (KNN) dan *Random Forest* untuk Klasifikasi Sentimen Komentar di TikTok (Studi Kasus: Performa ONIC Esports pada turnamen M7 *Mobile Legends: Bang Bang* (MLBB))

Oleh

ANGGIT PRATIKNYO WIDYAGIRI

22106010075

Performa ONIC Esports pada turnamen M7 *Mobile Legends: Bang Bang* (MLBB) memunculkan beragam opini pengguna TikTok sehingga diperlukan analisis sentimen untuk mengelompokkan komentar menjadi positif, negatif, dan netral. Penelitian ini bertujuan untuk mengetahui tahapan klasifikasi sentimen menggunakan metode *k-Nearest Neighbor* (kNN) dan *Random Forest*, menganalisis hasil klasifikasi sentimen, serta membandingkan performa kedua metode. Penelitian menggunakan 5.203 komentar berbahasa Indonesia yang diperoleh melalui *web scraping*. Data diproses melalui pelabelan, *preprocessing*, ekstraksi fitur *Term Frequency-Inverse Document Frequency* (TF-IDF), kemudian diklasifikasikan menggunakan kNN dan *Random Forest* dengan *Grid Search* dan *5-Fold Cross Validation*. Evaluasi model menggunakan confusion matrix, *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa *Random Forest* memberikan performa terbaik dengan akurasi 83,18%, sedangkan kNN dengan jarak *Euclidean* dan *Manhattan* masing-masing memperoleh akurasi 79,49% dan 64,82%. Dengan demikian, *Random Forest* lebih baik dalam mengklasifikasikan sentimen komentar TikTok mengenai performa ONIC Esports pada turnamen M7 MLBB.

Kata kunci: analisis sentimen, TikTok, *k-Nearest Neighbor*, *Random Forest*, TF-IDF, *Mobile Legends: Bang Bang*

ABSTRACT**Perbandingan Metode *K-Nearest Neighbor* (KNN) dan *Random Forest* untuk Klasifikasi Sentimen Komentar di TikTok (Studi Kasus: Performa ONIC Esports pada turnamen M7 Mobile Legend: Bang Bang (MLBB))**

by

ANGGIT PRATIKNYO WIDYAGIRI

22106010075

The performance of ONIC Esports in the M7 Mobile Legends: Bang Bang (MLBB) tournament generated various opinions among TikTok users, making sentiment analysis necessary to classify comments into positive, negative, and neutral categories. This study aims to identify the stages of sentiment classification using the k-Nearest Neighbor (kNN) and Random Forest methods, analyze the sentiment classification result, and compare the performance of both methods. The study used 5,203 Indonesian language comments collected through web scraping. The data were processed through labelling, preprocessing, and Term Frequency-Inverse Document Frequency (TF-IDF) feature extraction before being classified using kNN and Random Forest with Grid Search and 5-Fold Cross Validation. Model performance was evaluated using a confusion matrix, accuracy, precision, recall, and F1-score. The result showed that Random Forest achieved the best performance with an accuracy of 83,18%, while kNN using Euclidean and Manhattan distances achieved accuracies of 79.49% and 64.82%, respectively. Therefore, Random Forest outperformed kNN in classifying TikTok comments related to the performance of ONIC Esports in the M7 MLBB tournament.

Keywords: sentiment analysis, TikTok, k-Nearest Neighbor, Random Forest, TF-IDF, Mobile Legends: Bang Bang

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi dan komunikasi telah mendorong perubahan besar dalam pola interaksi masyarakat, terutama melalui media sosial. Media sosial tidak lagi hanya digunakan sebagai sarana komunikasi dan hiburan, tetapi juga menjadi ruang publik digital tempat masyarakat menyampaikan pendapat, kritik, dukungan, serta penilaian terhadap berbagai isu yang sedang berkembang. Tingginya aktivitas pengguna media sosial menghasilkan volume data teks yang sangat besar dan terus bertambah setiap waktu, sehingga analisis secara manual menjadi kurang efektif dan membutuhkan waktu yang lama. Kondisi tersebut mendorong berkembangnya analisis sentimen sebagai salah satu pendekatan dalam bidang *text mining* untuk mengidentifikasi opini publik secara otomatis (Xu et al., 2022).

Analisis sentimen merupakan teknik dalam *machine learning* yang digunakan untuk mengidentifikasi dan mengelompokkan opini masyarakat ke dalam kategori tertentu berdasarkan kecenderungan emosinya, seperti positif, negatif, dan netral (Rizkiyanto et al., 2024). Pendekatan ini memungkinkan data teks yang sebelumnya tidak terstruktur dapat diolah menjadi informasi yang lebih sistematis sehingga dapat membantu memahami kecenderungan opini masyarakat terhadap suatu topik tertentu. Selain itu, perkembangan *text mining* dan *machine learning* juga memberikan kemampuan bagi sistem untuk mengenali pola dari data teks secara otomatis melalui proses pembelajaran dari data yang tersedia (Allahyari et al., 2017). Dengan kemampuan tersebut, analisis sentimen menjadi metode yang relevan untuk mendukung proses analisis opini publik pada media sosial secara lebih cepat, efisien, dan terukur.

Salah satu media sosial yang memiliki tingkat interaksi tinggi saat ini adalah TikTok. TikTok berkembang menjadi platform digital yang tidak hanya digunakan

untuk berbagi video hiburan, tetapi juga menjadi ruang diskusi publik terhadap berbagai isu, seperti politik, olahraga, hiburan, pendidikan, dan perkembangan teknologi. Berdasarkan laporan *DataReportal* yang disusun oleh Simon Kemp (2025), TikTok memiliki 180 juta pengguna berusia 18 tahun ke atas di Indonesia pada akhir 2025 berdasarkan data audiens iklan TikTok. Jumlah tersebut menunjukkan bahwa TikTok memiliki jangkauan yang luas di Indonesia, dengan jangkauan iklan yang setara dengan 78,2% dari pengguna internet Indonesia (Kemp, 2025). Selain itu, fasilitas komentar pada TikTok memungkinkan pengguna memberikan tanggapan secara langsung terhadap suatu konten sehingga menghasilkan data teks yang dapat dimanfaatkan sebagai sumber informasi opini publik. Oleh karena itu, komentar pada platform TikTok memiliki potensi yang besar untuk dianalisis menggunakan pendekatan analisis sentimen.

Salah satu topik yang banyak diperbincangkan masyarakat di media sosial, khususnya TikTok, adalah perkembangan esports. Esports merupakan bentuk kompetisi olahraga elektronik yang melibatkan permainan video secara kompetitif dan profesional. Dalam beberapa tahun terakhir, esports mengalami perkembangan yang sangat pesat di Indonesia, terutama pada permainan *Mobile Legends: Bang Bang* (MLBB) yang menjadi salah satu permainan paling populer di kalangan masyarakat, khususnya remaja dan dewasa muda (Jiwandono & Purwandi, 2020). Popularitas MLBB tidak hanya terlihat dari jumlah pemain yang aktif, tetapi juga dari tingginya antusiasme masyarakat terhadap turnamen dan kompetisi internasional yang diselenggarakan secara rutin.

Salah satu turnamen internasional terbesar pada permainan tersebut adalah *M7 World Championship* yang diselenggarakan pada bulan Januari 2026. Turnamen ini menghadirkan berbagai tim profesional dari banyak negara, termasuk ONIC Esports sebagai salah satu perwakilan tim Indonesia (MPL Indonesia, 2026). Dalam pelaksanaannya, performa ONIC Esports pada turnamen M7 menjadi perhatian besar masyarakat karena hasil pertandingan yang diperoleh dinilai belum memenuhi ekspektasi sebagian pendukungnya (Jaelani, 2026). Kondisi tersebut memunculkan berbagai komentar dari pengguna TikTok yang berisi dukungan,

kritik, kekecewaan, hingga penilaian terhadap performa tim. Komentar-komentar tersebut mencerminkan opini publik yang beragam dan dapat dikategorikan ke dalam sentimen positif, negatif, dan netral. Banyaknya interaksi dan opini yang muncul menjadikan komentar TikTok mengenai performa ONIC Esports pada turnamen M7 sebagai objek penelitian yang menarik untuk dianalisis menggunakan pendekatan analisis sentimen.

Komentar pengguna TikTok memiliki karakteristik yang berbeda dibandingkan dengan teks formal pada umumnya karena cenderung menggunakan bahasa yang singkat, tidak baku, kontekstual, serta mengandung unsur emosional yang tinggi. Karakteristik tersebut menyebabkan data komentar media sosial menjadi tidak terstruktur dan proses klasifikasi sentimen menjadi lebih kompleks untuk dianalisis secara manual. Oleh karena itu, diperlukan pendekatan komputasi yang mampu mengenali pola data teks secara otomatis, salah satunya melalui *machine learning*. *Machine learning* merupakan cabang dari kecerdasan buatan yang memungkinkan sistem komputer mempelajari pola dari data dan melakukan prediksi atau klasifikasi secara otomatis tanpa harus diprogram secara eksplisit. Dalam bidang analisis sentimen, *machine learning* digunakan untuk mengidentifikasi pola kata, hubungan antar fitur, seperti keterkaitan antar kata yang muncul dalam suatu teks, serta kecenderungan sentimen dari teks tersebut berdasarkan data pelatihan yang tersedia (Allahyari et al., 2017). Pendekatan ini dinilai efektif untuk mengolah data media sosial yang memiliki volume besar, bersifat dinamis, dan tidak terstruktur.

Proses klasifikasi sentimen pada data komentar media sosial tidak dapat dilakukan secara langsung karena data teks yang diperoleh umumnya masih bersifat mentah dan tidak terstruktur (Lin & Nuha, 2023). Oleh karena itu, diperlukan beberapa tahapan pengolahan data sebelum proses klasifikasi dilakukan. Tahapan tersebut meliputi pengumpulan data komentar, *preprocessing* teks, transformasi data teks menjadi representasi numerik, hingga proses pelatihan dan pengujian model klasifikasi menggunakan metode *machine learning*. Tahapan *preprocessing*

teks umumnya mencakup *cleaning*, *case folding*, *tokenizing*, *normalization*, *stemming*, dan *stopword removal*.

Setelah melalui *preprocessing*, data teks kemudian diubah menjadi bentuk numerik menggunakan teknik ekstraksi fitur seperti *Term Frequency-Inverse Document Frequency* (TF-IDF) agar dapat diproses oleh metode klasifikasi (Ahuja et al., 2019). Proses pembobotan TF-IDF merupakan salah satu penerapan konsep matematis dalam *text mining* yang digunakan untuk mengukur tingkat kepentingan suatu kata dalam dokumen. Selanjutnya, hasil pembobotan tersebut digunakan sebagai dasar dalam proses klasifikasi sentimen menggunakan metode *machine learning*. Seluruh tahapan tersebut menjadi bagian penting dalam analisis sentimen karena kualitas pengolahan data sangat memengaruhi hasil klasifikasi yang dihasilkan. Dengan adanya tahapan klasifikasi yang sistematis, proses identifikasi sentimen pada komentar TikTok dapat dilakukan secara lebih optimal dan terukur.

Salah satu metode *machine learning* yang sering digunakan dalam klasifikasi teks adalah *k-Nearest Neighbor* (kNN). Metode KNN bekerja dengan cara menentukan kelas suatu data berdasarkan kedekatan jarak dengan sejumlah data latih terdekat. Proses klasifikasi dilakukan dengan memilih kelas mayoritas dari tetangga terdekat yang dimiliki data tersebut (Cover & Hart, 1967). Metode ini memiliki implementasi yang relatif sederhana dan efektif digunakan dalam klasifikasi teks pendek karena mampu mengenali kemiripan pola antar data. Selain kNN, metode *Random Forest* juga banyak digunakan dalam proses klasifikasi sentimen. *Random Forest* merupakan metode *ensemble learning* yang bekerja dengan membangun sejumlah pohon keputusan (*Decision Tree*) dan menggabungkan hasil prediksi dari setiap pohon untuk memperoleh hasil klasifikasi yang lebih akurat dan stabil. Metode *Random Forest* memiliki kemampuan yang baik dalam meningkatkan akurasi klasifikasi serta mengurangi risiko *overfitting* pada proses prediksi (Breiman, 2001).

Pemilihan metode *k-Nearest Neighbor* (kNN) dan *Random Forest* pada skripsi ini didasarkan pada karakteristik kedua metode yang memiliki pendekatan klasifikasi berbeda. kNN merupakan metode berbasis kedekatan jarak antar data,

sedangkan *Random Forest* merupakan metode *ensemble learning* berbasis *decision tree*. Perbedaan mekanisme tersebut memungkinkan dilakukan perbandingan performa klasifikasi pada data komentar TikTok yang bersifat dinamis dan tidak terstruktur. Selain itu, kedua metode memiliki implementasi yang relatif sederhana dan efisien serta banyak digunakan dalam penelitian klasifikasi teks sehingga relevan untuk diterapkan pada skripsi ini (Yanti et al., 2023).

Berdasarkan keterbatasan penelitian terdahulu, skripsi ini berfokus pada analisis sentimen komentar TikTok mengenai performa ONIC Esports pada turnamen M7 MLBB. Penelitian ini mencakup tahapan pengumpulan data, *preprocessing*, ekstraksi fitur TF-IDF, klasifikasi menggunakan metode *k-Nearest Neighbor* (kNN) dan *Random Forest*, serta evaluasi model berdasarkan akurasi, presisi, *recall*, dan F1-score. Hasil penelitian diharapkan dapat memberikan gambaran empiris mengenai metode yang lebih efektif dalam klasifikasi sentimen komentar TikTok pada konteks Esports.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana langkah klasifikasi sentimen menggunakan metode *k-Nearest Neighbor* (kNN) dan *Random Forest*?
2. Bagaimana hasil klasifikasi sentimen komentar di TikTok mengenai performa ONIC Esports pada turnamen M7 MLBB menggunakan metode *k-Nearest Neighbor* (kNN) dan *Random Forest*?
3. Bagaimana perbandingan antara metode *k-Nearest Neighbor* (kNN) dan *Random Forest* dalam mengklasifikasikan sentimen komentar di TikTok mengenai performa ONIC Esports pada turnamen M7 MLBB?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang diuraikan, tujuan dari skripsi ini adalah sebagai berikut:

1. Mengetahui langkah-langkah klasifikasi sentimen komentar di TikTok menggunakan metode *k-Nearest Neighbor* (kNN) dan *Random Forest*.
2. Mengetahui dan menganalisis hasil klasifikasi sentimen komentar di TikTok mengenai performa ONIC Esports pada turnamen M7 MLBB menggunakan metode *k-Nearest Neighbor* (kNN) dan *Random Forest*.
3. Membandingkan performa metode *k-Nearest Neighbor* (kNN) dan *Random Forest* untuk menentukan metode yang lebih baik dalam mengklasifikasikan sentimen komentar di TikTok mengenai performa ONIC Esports pada turnamen M7 MLBB.

1.4 Batasan Masalah

Berdasarkan tujuan masalah yang telah dijelaskan sebelumnya, maka batasan masalah dalam skripsi ini sebagai berikut:

1. Data yang digunakan berupa komentar berbahasa Indonesia pada platform TikTok mengenai performa ONIC Esports pada turnamen M7 MLBB.
2. Klasifikasi sentimen dibatasi pada tiga kategori, yaitu positif, negatif, dan netral.
3. Penelitian tidak membahas analisis konteks mendalam seperti sarkasme atau ironi dalam teks.
4. Penelitian membandingkan kinerja metode kNN dan *Random Forest* berdasarkan nilai akurasi, presisi, *recall*, dan *F1-Score*.
5. Penelitian menggunakan Python versi 3.14.6 dengan Visual Studio Code sebagai aplikasi pemograman serta library Pandas, NumPy, dan Sckit-learn untuk pengolahan data, pemodelan, dan evaluasi

1.5 Manfaat Penelitian

Adapun manfaat yang diharapkan dari penelitian ini dibagi menjadi dua, yaitu manfaat teoritis dan manfaat praktis sebagai berikut:

1. Manfaat Teoritis

- a. Memberikan kontribusi terhadap pengembangan kajian analisis sentimen melalui penerapan metode *k-Nearest Neighbor* (kNN) dan *Random Forest* pada klasifikasi komentar di TikTok.
- b. Menjadi referensi bagi penelitian selanjutnya yang membahas perbandingan metode klasifikasi dalam analisis sentimen.

2. Manfaat Praktis

- a. Memberikan pengalaman bagi penulis dalam pengolahan data teks dan penerapan metode *k-Nearest Neighbor* (kNN) dan *Random Forest* untuk klasifikasi sentimen.
- b. Memberikan informasi mengenai metode yang memiliki kinerja lebih baik dalam klasifikasi sentimen komentar di TikTok mengenai performa ONIC Esports pada turnamen M7 MLBB.
- c. Memberikan gambaran mengenai respons masyarakat terhadap performa ONIC Esports pada turnamen M7 MLBB berdasarkan komentar yang dianalisis.

1.6 Tinjauan Pustaka

Pustaka utama dalam skripsi ini adalah jurnal yang ditulis oleh Thomas M. Cover dan Peter E. Hart (1967). Penelitian ini menjadi dasar teoritis metode kNN karena membahas klasifikasi berbasis kedekatan jarak, pendekatan non-parametrik, serta pengaruh nilai k terhadap hasil klasifikasi. Perbedaannya, penelitian ini lebih menekankan konsep dasar dan mekanisme kerja algoritma kNN. Pustaka utama lainnya adalah jurnal Leo Breiman (2001). Penelitian ini menjelaskan konsep dasar *Random Forest* sebagai metode *ensemble learning* melalui teknik *bagging* dan pemilihan fitur acak pada pohon keputusan. Perbedaannya terletak pada fokus penelitian yang menekankan peningkatan akurasi dan pengurangan *overfitting* pada model klasifikasi.

Tinjauan pustaka berikutnya adalah penelitian analisis sentimen yang ditulis oleh Rizkiyanto et al. (2024). Penelitian ini menerapkan TF-IDF dan *Random Forest* untuk klasifikasi sentimen ulasan aplikasi PLN Mobile. Hasil penelitian

menunjukkan model terbaik diperoleh menggunakan *stemming* dan *unigram* TF-IDF dengan *F1-Score* sebesar 93,14%. Perbedaannya, penelitian ini berfokus pada ulasan aplikasi PLN Mobile dengan metode *Random Forest* pada analisis sentimen berbasis teks.

Penelitian lainnya dilakukan oleh Supriyadi dan Makatita (2025) mengenai analisis sentimen pengguna TikTok menggunakan metode *IndoBERT*. Hasil penelitian menunjukkan bahwa metode *IndoBERT* mampu memberikan performa klasifikasi yang baik pada data komentar TikTok. Perbedaannya, penelitian tersebut menggunakan metode *IndoBERT* dan membahas topik adaptasi QRIS. Kemudian, terdapat jurnal yang ditulis oleh Leif E. Peterson (2009). Penelitian ini membahas konsep kNN sebagai metode *supervised learning* dan *instance-based learning* menggunakan *Euclidean Distance* dalam menentukan kelas data. Perbedaannya, penelitian ini lebih menekankan prinsip kerja dan karakteristik algoritma kNN secara konseptual.

Penelitian ini berfokus pada analisis sentimen komentar TikTok menggunakan perbandingan metode kNN dan *Random Forest* untuk mengetahui metode dengan performa klasifikasi terbaik pada data komentar terkait performa ONIC Esport di turnamen M7 MLBB.

Tabel 1.1 Perbedaan dengan penelitian sebelumnya

No.	Nama Penulis	Hasil	Persamaan	Perbedaan
1.	Thomas M. Cover dan Peter E. Hart (1967)	Memperkenalkan metode kNN untuk klasifikasi berbasis tetangga terdekat.	Menggunakan metode kNN.	Fokus pada dasar teori dan konsep klasifikasi kNN berbasis kedekatan jarak.
2.	Leo Breiman (2001)	Memperkenalkan metode <i>Random Forest</i> sebagai metode <i>ensemble learning</i> .	Menggunakan metode <i>Random Forest</i> .	Membahas konsep <i>Random Forest</i> untuk meningkatkan akurasi dan mengurangi overfitting.
3.	Rizkiyanto et al. (2024)	<i>Random Forest</i> dengan TF-IDF memperoleh F1-Score 93,14%.	Menggunakan TF-IDF, <i>Random Forest</i> , dan analisis sentimen.	Menggunakan TF-IDF dan <i>Random Forest</i> pada ulasan PLN Mobile dengan F1-Score 93,14%.
4.	Leif E. Peterson (2009)	Menjelaskan konsep dasar dan cara kerja kNN.	Menggunakan metode kNN.	Menjelaskan prinsip kerja kNN dan Euclidean Distance pada klasifikasi data.
5.	Anggit Pratiknyo W. (2026)	<i>Random Forest</i> memiliki performa yang lebih baik dengan nilai <i>accuracy</i> sebesar 83,18%.	Menggunakan TF-IDF, <i>Random Forest</i> , dan analisis sentimen.	Membandingkan kNN dan <i>Random Forest</i> pada komentar TikTok ONIC Esport.

1.7 Sistematika Penulisan

Sistematika penulisan dalam skripsi ini terdiri dari enam bab sebagai berikut:

BAB I : PENDAHULUAN

Pada Bab I dibahas mengenai latar belakang dan rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, tinjauan pustaka dan sistematika penulisan sebagai penutup bab.

BAB II : LANDASAN TEORI

Dalam Bab II dijelaskan mengenai teori-teori yang menjadi dasar dalam penelitian klasifikasi sentimen komentar TikTok, meliputi vektor, matriks, *data mining*, *text mining*, analisis sentimen, *preprocessing*, representasi teks, *vector space model*, pembobotan TF-IDF, *K-Fold Cross Validation*, *Grid Search*, dan *Confusion Matrix*. Kemudian dibahas teori *machine learning*, *supervised learning*, jarak, dan *Decision Tree* yang menjadi teori dasar dari metode yang digunakan, yaitu kNN dan *Random Forest*.

BAB III : METODE PENELITIAN

Bab ini membahas metode penelitian yang digunakan dalam penelitian klasifikasi sentimen komentar TikTok mengenai performa ONIC Esports pada turnamen M7 MLBB, meliputi jenis penelitian, metode pengumpulan data, variabel penelitian, metode penelitian, metode analisis data, serta alur penelitian yang digunakan dalam proses klasifikasi sentimen menggunakan metode kNN dan *Random Forest*.

BAB IV : PEMBAHASAN

Bab ini membahas metode kNN dan *Random Forest* yang digunakan dalam penelitian klasifikasi sentimen komentar TikTok, meliputi pengertian dan konsep dasar, parameter, cara kerja, pengujian model, dan contoh perhitungan manual masing-masing metode.

BAB V : STUDI KASUS

Bab ini menerapkan metode kNN dan *Random Forest* pada kasus klasifikasi sentimen komentar TikTok mengenai performa ONIC Esports pada turnamen M7 MLBB, meliputi tahapan pengolahan data komentar, pelabelan sentimen, *preprocessing* teks, pembobotan TF-IDF, pembagian data latih dan data uji, proses klasifikasi menggunakan metode kNN dan *Random Forest*, serta

hasil klasifikasi dan evaluasi kinerja kedua metode. Selanjutnya, dijelaskan mengenai perbandingan hasil klasifikasi menggunakan *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix* untuk mengetahui metode yang memberikan kinerja lebih baik dalam klasifikasi sentimen komentar TikTok.

BAB VI : PENUTUP

Bab ini berisi kesimpulan dan saran yang diperoleh dari hasil penelitian klasifikasi sentimen komentar TikTok mengenai performa ONIC Esports pada turnamen M7 MLBB menggunakan metode kNN dan *Random Forest*. Kesimpulan merangkum hasil penelitian berdasarkan tujuan yang telah ditetapkan, sedangkan saran disampaikan sebagai bahan pertimbangan bagi penelitian selanjutnya.

BAB VI

PENUTUP

6.1 Kesimpulan

Proses klasifikasi sentimen dilakukan melalui beberapa tahapan, yaitu pengumpulan data komentar TikTok, pelabelan sentimen, *preprocessing* yang meliputi *cleaning*, *case folding*, tokenisasi, normalisasi, *stopword removal*, dan *stemming*, pembagian data menjadi data latih dan data uji, ekstraksi fitur menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF), serta pembentukan dan pengujian model menggunakan metode kNN dan *Random Forest*. Pada metode kNN, parameter terbaik diperoleh menggunakan *Grid Search* dengan *5-Fold Cross Validation*, yaitu nilai $k = 19$. Sementara itu, pada metode *Random Forest*, diperoleh kombinasi parameter terbaik diperoleh dengan menggunakan 200 pohon keputusan (*n_estimators*), tanpa membatasi kedalaman maksimum pohon (*max_depth = None*), setiap node dibagi apabila memiliki sedikitnya 5 sampel (*min_samples_split = 5*), dan setiap node daun berisi minimal 1 sampel (*min_samples_leaf*).

Hasil klasifikasi sentimen menunjukkan bahwa model kNN dengan metrik jarak *Euclidean* memperoleh nilai *accuracy* sebesar 79,49%, *precision* sebesar 68,7%, *recall* sebesar 65,43%, dan *F1-score* sebesar 65,98%. Sementara itu, kNN dengan metrik jarak *Manhattan* memperoleh nilai *accuracy* sebesar 64,82%, *precision* sebesar 60,95%, *recall* sebesar 56,57%, dan *F1-score* sebesar 56,04%. Adapun metode *Random Forest* memperoleh nilai *accuracy* sebesar 83,18%, *precision* sebesar 73,64%, *recall* sebesar 70,24%, dan *F1-score* sebesar 71,24%.

Berdasarkan hasil perbandingan kinerja, metode *Random Forest* memberikan performa terbaik dalam mengklasifikasikan sentimen komentar TikTok mengenai performa ONIC Esports pada turnamen M7 MLBB. Selain memiliki nilai *accuracy*, *precision*, *recall*, dan *F1-score* yang lebih tinggi dibandingkan metode kNN, *Random Forest* juga menunjukkan kemampuan

klasifikasi yang lebih baik pada kelas sentimen positif dan negatif. Sementara itu, metode kNN dengan metrik jarak *Euclidean* menghasilkan performa yang lebih baik dibandingkan kNN dengan metrik jarak *Manhattan*. Dengan demikian, *Random Forest* merupakan metode yang paling efektif untuk klasifikasi sentimen pada dataset yang digunakan dalam penelitian.

6.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat dipertimbangkan untuk penelitian selanjutnya, yaitu sebagai berikut.

1. Penelitian selanjutnya disarankan menggunakan jumlah data yang lebih besar serta berasal dari berbagai sumber atau platform media sosial agar model yang dihasilkan memiliki kemampuan generalisasi yang lebih baik
2. Proses ekstraksi fitur dapat dikembangkan dengan menggunakan metode lain, seperti *Word2Vec*, *FastText*, atau *BERT embeddings*, untuk mengetahui pengaruh representasi fitur terhadap kinerja model klasifikasi sentimen.
3. Penelitian selanjutnya juga dapat melakukan pengembangan pada tahap *preprocessing*, khususnya dalam penyusunan kamus normalisasi kata tidak baku (*slang*) dan penanganan bahasa campuran, sehingga kualitas data yang digunakan dalam proses klasifikasi dapat ditingkatkan.

DAFTAR PUSTAKA

- Aggarwal, C. C. (2018). *Neural Networks and Deep Learning* (1st ed.). Springer International Publishing. <https://doi.org/10.1007/978-3-319-94463-0>
- Ahuja, R., Chug, A., Kohli, S., Gupta, S., & Ahuja, P. (2019). The impact of features extraction on the sentiment analysis. *Procedia Computer Science*, 152, 341–348. <https://doi.org/10.1016/j.procs.2019.05.008>
- Aida, Hermina, D., & Norlaila. (2025). Jenis Data Penelitian Kuantitatif (Korelasional, Komparatif, dan Eksperimen). *AL-MANBA, Jurnal Ilmiah Keislaman Dan Kemasyarakatan*, 10(1), 1–10.
- Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017, July 28). A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques. *Proceedings of KDD Bigdas 2017*. <http://arxiv.org/abs/1707.02919>
- Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization Yoshua Bengio. In *Journal of Machine Learning Research* (Vol. 13). <http://scikit-learn.sourceforge.net>.
- Biau, G., & Scornet, E. (2015). A Random Forest Guided Tour. *TEST*, 25(2). <http://arxiv.org/abs/1511.05741>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Budiman, I., Faisal, M. R., & Nugraha, D. T. (2020). Studi Ekstraksi Fitur Berbasis Vektor Word2Vec pada Pembentukan Fitur Berdimensi Rendah. *Jurnal Komputasi*, 8(1). <https://doi.org/10.23960/komputasi.v8i1.2517>
- Chrisinta, D., & Simarmata, J. E. (2024). Eksplorasi Teknik Web Scraping pada Data Mining: Pendekatan Pencarian Data Berbasis Python. *Faktor Exacta*, 17(1). <https://doi.org/10.30998/faktorexacta.v17i1.22393>
- Cover, T. M., & Hart, P. E. (1967). Approximate formulas for the information transmitted by a discrete communication channel. In *IEEE TRANSACTIONS ON INFORMATION THEORY* (Vol. 24, Number 1).
- Daulay, R. S. (2024). Analisis Kritis dan Pengembangan Algoritma K-Nearest Neighbor (KNN): Sebuah Tinjauan Literatur. *Jurnal Pendidikan Sains Dan Komputer*, 4(02), 131–141. <https://doi.org/10.47709/jpsk.v4i02.5055>
- Fredriksson, T., Mattos, D. I., Bosch, J., & Olsson, H. H. (2020). Data Labeling: An Empirical Investigation into Industrial Challenges and Mitigation

- Strategies. In *Product-Focused Software Process Improvement* (pp. 202–216). Springer. https://doi.org/10.1007/978-3-030-64148-1_13
- Fullah, A., Septian, M. A., Putra, M. R. H., Fadiatama, M. I., Al Farizi, V., Puspitasari, N., & Hairah, U. (2025). Analisis Sentimen Isu Penyalahgunaan Data pada Layanan Pinjaman Online Menggunakan Support Vector Machine di Platform X. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(3S1). <https://doi.org/10.23960/jitet.v13i3S1.7976>
- Guo, Z., Liu, G., Liu, W., Ni, Z., & Jiang, F. (2026). GWKNN: an enhanced k-nearest neighbor algorithm with G metric reconstruction and Grey Wolf Optimizer. *Scientific Reports*, 16(1). <https://doi.org/10.1038/s41598-026-41851-2>
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)* (3rd ed.). Morgan Kaufmann.
- Hotho, A., Nürnberger, A., & Paaß, G. (2005). A Brief Survey of Text Mining. *Journal for Language Technology and Computational Linguistics*, 20, 19–62. <https://doi.org/10.21248/jlcl.20.2005.68>
- Jaelani, D. (2026, January 22). *Hasil Knockout M7 ONIC Esports vs TLPH: Langkah Si Landak Terhenti*. <https://www.ggwp.id/esports/moba/hasil-knockout-m7-onic-esports-vs-tlph-00-zrzwz-kh0lll>
- Jiwandono, H. P., & Purwandi, E. R. (2020). Liga Seluler: Pergeseran Olahraga Elektronik ke Peran Seluler dalam Kompetisi Mobile Legends: Bang Bang di Indonesia. *Jurnal Media Dan Komunikasi*, 1(1), 4–18.
- Kemp, S. (2025, November 5). *Digital 2026: Indonesia*. DataReportal. DataReportal - Digital 2026: Indonesia
- Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95)*, 1137–1145. <http://robotics.stanford.edu/~ronnyk>
- Lin, C. H., & Nuha, U. (2023). Sentiment analysis of Indonesian datasets based on a hybrid deep-learning strategy. *Journal of Big Data*, 10(1). <https://doi.org/10.1186/s40537-023-00782-9>
- M, A., & E, S. (2019). A Study of Feature Extraction techniques for Sentiment Analysis. In *Emerging Technologies in Data Mining and Information Security* (Vol. 814, pp. 475–486). Springer. https://doi.org/10.1007/978-981-13-1501-5_41

- Mienye, I. D., & Jere, N. (2024). A Survey of Decision Trees: Concepts, Algorithms, and Applications. *IEEE Access*, 12, 86716–86727. <https://doi.org/10.1109/ACCESS.2024.3416838>
- Mishra, K. M., & Huhtala, K. (2020). Condition Monitoring of Elevator Systems using Deep Neural Network. *International Conference on Operations Research and Enterprise Systems*, 381–387. <https://doi.org/10.5220/0009348803810387>
- MPL Indonesia. (2026, January 10). *M7 World Championship*. <https://id-mpl.com/en/m7>
- Mudarakola, L. P., Gatla, R. K., Raju, A. S. N., Jaffar, A. Y., Alzahrani, A., & Dessalegn, A. A. (2025). Multi stage sentiment analysis for product reviews on Twitter using optimized machine learning algorithm. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-23451-8>
- Peterson, L. (2009). K-nearest neighbor. *Scholarpedia*, 4(2), 1883. <https://doi.org/10.4249/scholarpedia.1883>
- Quinlan, J. R. (1986). Induction of Decision Trees. In *Machine Learning* (Vol. 1).
- Rifaldi, D., Abdul Fadlil, & Herman. (2023). Teknik Preprocessing Pada Text Mining Menggunakan Data Tweet “Mental Health.” *Decode: Jurnal Pendidikan Teknologi Informasi*, 3(2), 161–171. <https://doi.org/10.51454/decode.v3i2.131>
- Rizkiyanto, M. D., Purbolaksono, M. D., & Astuti, W. (2024). Sentiment Analysis Classification on PLN Mobile Application Reviews using Random Forest Method and TF-IDF Feature Extraction. *INTEK: Jurnal Penelitian*, 11(1), 37–43. <https://doi.org/10.31963/intek.v11i1.4774>
- Salsabila, Martha, S., & Andani, W. (2024). Komparasi algoritma K-Nearest Neighbor dengan Euclidean Distance dan Manhattan Distance untuk klasifikasi stunting balita (Studi Kasus : Puskesmas Kelurahan Parit Mayor). *Buletin Ilmiah Math. Stat. Dan Terapannya (Bimaster)*, 13, 285–292.
- Salton, G., Wong, A., & Yang, C. S. (1975). A Vector Space Model for Automatic Indexing. *Communications of the ACM*, 18(11), 613–620.
- Scikit-learn. (2024). *TfidfVectorizer*. Scikit-Learn. https://scikit-learn.org/1.5/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html
- Scornet, E., Biau, G., & Vert, J. P. (2015). Consistency of random forests. *Annals of Statistics*, 43(4), 1716–1741. <https://doi.org/10.1214/15-AOS1321>

- Sinoara, R. A., Antunes, J., & Rezende, S. O. (2017). Text mining and semantics: a systematic mapping study. *Journal of the Brazilian Computer Society*, 23(1). <https://doi.org/10.1186/s13173-017-0058-7>
- Strang, G. (2009). *Introduction To Linear Algebra* (4th ed.). Wellesley-Cambridge Press.
- Sumarni, Ondeng, S., & Rahman, U. (2025). Ragam Variabel dan definisi Operasional Penelitian Pendidikan: Implikasi Metodologis Pendidikan Islam dan Pendidikan Matematika. *Jurnal Pendidikan, Penciptaan Seni Dan Budaya*, 1(2), 80–88. <https://doi.org/10.62710/jppsbs4m3n888>
- Suroyya, C. F., Mahmudah, I., & Fatimah, S. (2022). Konsep Dasar Metodologi Penelitian Pada Bidang Pendidikan Dasar. *Al-Madrasah: Jurnal Pendidikan Madrasah Ibtidaiyah*, 6(4), 1255. <https://doi.org/10.35931/am.v6i4.1187>
- Susianti, O. M., & Srifariyati. (2024). Perumusan Variabel Dan Indikator Dalam Penelitian Kuantitatif Kependidikan. *Jurnal Pendidikan Rokania*, 9, 18–30. <https://doi.org/10.37728/jpr.v9i1.1066>
- Vega-Huerta, H., Pajuelo-Leon, J., De-la-Cruz-VdV, P., Calderón, D., Maquen-Niño, G. L. E., Rios-Castillo, M. E., Camara-Figueroa, A., Gil-Calvo, R., Guerra-Grados, L., & Benito-Pacheco, O. (2025). K-Nearest Neighbors Model to Optimize Data Classification According to the Water Quality Index of the Upper Basin of the City of Huarmey. *Applied Sciences (Switzerland)*, 15(18). <https://doi.org/10.3390/app151810202>
- Widyaningsih, Y., Arum, G. P., & Prawira, K. (2021). Application of k-fold cross validation in determining the best negative binomial regression model. *Barekeng: Jurnal Ilmu Matematika Dan Terapan*, 15(2), 315–322. <https://doi.org/10.30598/barekengvol15iss2pp315-322>
- Xu, Q. A., Chang, V., & Jayne, C. (2022). A systematic review of social media-based sentiment analysis: Emerging trends and challenges. *Decision Analytics Journal*, 3, 100073. <https://doi.org/10.1016/j.dajour.2022.100073>
- Yanti, C. P., Eva Agustini, N. W., Sri Rahayu Ginantra, N. L. W., & Putri Wulandari, D. A. (2023). Perbandingan Metode K-NN Dan Metode Random Forest Untuk Analisis Sentimen pada Tweet Isu Minyak Goreng di Indonesia. *Jurnal Media Informatika Budidarma*, 7(2), 756. <https://doi.org/10.30865/mib.v7i2.5900>
- Yıldırım, S. (2020, February 11). *Decision Tree and Random Forest Explained*. Towards Data Science. <https://towardsdatascience.com/decision-tree-and-random-forest-explained-8d20ddabc9dd/>