

Analisis Sentimen Dampak Perkembangan *Artificial Intelligence* (AI) pada Media Sosial X/Twitter Menggunakan Metode Random Forest

Rizqi Wijanarko¹, Dian Eka Ratnawati², Putra Pandu Adikara³

Program Studi Teknologi Informasi, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹rizqiwijanarko@student.ub.ac.id, ²dian_ilkom@ub.ac.id, ³adikara.putra@ub.ac.id

Abstrak

Kecerdasan buatan atau *Artificial Intelligence* (AI) adalah salah satu inovasi teknologi yang saat ini sedang berkembang dengan pesat dan menarik perhatian. *Artificial Intelligence* adalah suatu program komputer yang dilengkapi dengan algoritma yang mampu memperoleh pengetahuan dari data yang diberikan serta mengaplikasikannya dalam berpikir dan bertindak selayaknya manusia. Dengan berkembangnya *Artificial Intelligence*, maka tidak dapat dipungkiri bahwa teknologi ini memberikan dampak positif maupun negatif terhadap kehidupan masyarakat. Banyak opini dari masyarakat yang timbul di media sosial terkait dampak perkembangan *Artificial Intelligence* di Indonesia salah satunya melalui media sosial X/Twitter. Salah satu cara untuk mengetahui persepsi masyarakat dapat dilakukan melalui analisis sentimen. Penelitian ini menggunakan metode Random Forest sebagai metode klasifikasi untuk analisis sentimen. Terdapat beberapa tahapan yang dilakukan, yaitu pengumpulan data menggunakan *web scraping*, *text preprocessing*, pembobotan kata menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF), klasifikasi menggunakan Random Forest, serta evaluasi kinerja menggunakan Confusion Matrix. Hasil pengujian menunjukkan bahwa dengan parameter jumlah *tree* 60 dan kedalaman *tree* 55 mendapatkan tingkat *accuracy* sebesar 0,94, *precision* sebesar 0,94, *recall* sebesar 0,94, dan *f1-score* sebesar 0,94.

Kata kunci: *analisis sentimen, x/twitter, artificial intelligence, klasifikasi, random forest*

Abstract

Artificial Intelligence (AI) is one of the rapidly developing and attention-grabbing technological innovations. It is a computer program equipped with algorithms capable of acquiring knowledge from provided data and applying it in thinking and acting similar to humans. With the advancement of Artificial Intelligence, it is undeniable that this technology has both positive and negative impacts on society. Many opinions from the public arise on social media regarding the impact of Artificial Intelligence development in Indonesia, one of which is through the social media platform X/Twitter. One way to understand public perception is through sentiment analysis. This research utilizes the Random Forest method as a classification technique for sentiment analysis. Several stages are involved, including data collection using web scraping, text preprocessing, word weighting using Term Frequency-Inverse Document Frequency (TF-IDF), classification using Random Forest, and performance evaluation using a Confusion Matrix. The test result shows that with the parameters number of tree 60 and tree depth 55, the accuracy level is 0,94, precision is 0,94, recall is 0,94, and f1-score is 0,94.

Keywords: *sentiment analysis, x/twitter, artificial intelligence, classification, random forest*

1. PENDAHULUAN

Teknologi merupakan suatu hal yang terus berkembang tanpa henti seiring berjalannya waktu. Dalam perkembangannya selalu melahirkan inovasi-inovasi baru yang dapat memberikan berbagai macam dampak dalam kehidupan. Kecerdasan buatan atau *Artificial Intelligence* (AI) merupakan salah satu teknologi

yang saat ini sedang menarik perhatian dan mengalami kemajuan yang signifikan. *Artificial Intelligence* adalah suatu program komputer yang dilengkapi dengan algoritma yang mampu memperoleh pengetahuan dari data yang diberikan serta mengaplikasikannya dalam berpikir dan bertindak selayaknya manusia (Rouhiainen, 2018).

Dengan berkembangnya *Artificial Intelligence*, maka tidak dapat dipungkiri bahwa teknologi ini memberikan dampak positif maupun negatif terhadap kehidupan masyarakat. Dampak positif dari penggunaan *Artificial Intelligence*, seperti meningkatkan efisiensi dan produktivitas, kemajuan dalam bidang kesehatan, meningkatkan keamanan, serta kemudahan dan kenyamanan dalam kehidupan sehari-hari (Anugrah, 2023). *Artificial Intelligence* di sisi lain juga merupakan ancaman bagi sumber daya manusia, salah satunya karena dengan adanya berbagai kecanggihan dalam bidang *Artificial Intelligence* pekerjaan manusia perlahan semakin tergantikan (Pakpahan, 2021). Banyak opini dari masyarakat yang timbul di media sosial terkait dampak perkembangan *Artificial Intelligence*.

Analisis sentimen ini diperlukan untuk mendapatkan pemahaman tentang bagaimana masyarakat merespons teknologi *Artificial Intelligence* dan juga untuk melihat kinerja dari metode Random Forest yang digunakan dalam melakukan klasifikasi sentimen terhadap dampak perkembangan *Artificial Intelligence*. Pembuatan *Word Cloud* dapat berguna untuk membantu mengidentifikasi kata-kata dominan yang muncul dalam data. Kata-kata tersebut dapat dianalisis untuk memahami konteks dan sentimen yang melekat pada setiap kata, sehingga memberikan wawasan yang lebih mendalam tentang persepsi masyarakat. Indonesia juga belum memiliki regulasi khusus terkait *Artificial Intelligence*, maka persepsi masyarakat terkait hal tersebut dapat membantu pemerintah atau pemangku kebijakan dalam merancang kebijakan dan strategi yang sesuai dengan kebutuhan dan aspirasi masyarakat.

Media sosial Twitter atau yang saat ini bernama X adalah suatu media yang banyak dipakai untuk menyampaikan suatu opini dalam *tweet*. Opini tersebut dapat digunakan sebagai sumber yang berpotensi untuk menghasilkan informasi yang berguna. Analisis yang tepat diperlukan dalam pemanfaatannya untuk menghasilkan informasi yang bermanfaat bagi berbagai pihak untuk pengambilan keputusan, analisis sentimen adalah salah satu metodenya (Mahardhika & Zuliarso, 2018). Analisis sentimen adalah cabang ilmu komputasi yang dapat dimanfaatkan untuk menganalisis sentimen, opini, sikap, evaluasi, emosi, dan penilaian dari pernyataan seseorang kepada berbagai entitas, seperti layanan, individu,

organisasi, produk, permasalahan, acara, atribut, dan topik yang terkait (Liu, 2012). Analisis sentimen dapat digunakan untuk menganalisis teks untuk menentukan apakah mengandung opini yang positif, negatif, atau netral. Terdapat beberapa metode yang bisa diterapkan dalam analisis sentimen, seperti Random Forest, Support Vector Machine, K-NN, Naive Bayes, serta berbagai metode yang lainnya.

Terdapat beberapa penelitian terdahulu yang berkaitan dengan analisis sentimen, seperti penelitian yang dilaksanakan Fitri, et al. (2020) membandingkan metode Support Vector Machine, Naive Bayes, serta Random Forest dalam analisis sentimen aplikasi Ruangguru. Hasil dari penelitian tersebut menunjukkan metode Random Forest memperoleh akurasi tertinggi yaitu 97,16%, sedangkan metode Naive Bayes serta Support Vector Machine masing-masing memperoleh akurasi yaitu 96,01% dan 94,16%. Yanti, et al. (2023) melakukan perbandingan antara metode dan K-NN dalam analisis sentimen terhadap *tweet* yang berkaitan dengan isu minyak goreng di Indonesia. Hasilnya menunjukkan bahwa rata-rata akurasi metode Random Forest lebih unggul yaitu 73,37%, dan pada metode K-NN yaitu 72,86%. Sudioanto, et al. (2023) menggunakan metode Support Vector Machine serta Support Vector Machine Random Forest untuk melakukan perbandingan dalam menganalisis sentimen Twitter terkait kaburnya *selebgram* Rachel Vennya dari karantina. Hasil dari penelitian tersebut memperlihatkan metode Random Forest memperoleh akurasi lebih tinggi yaitu yaitu 94%, sedangkan pada metode Support Vector Machine yaitu 93%. Berdasarkan beberapa penelitian terdahulu dapat disimpulkan bahwa Random Forest memiliki kemampuan dan tingkat akurasi baik dalam mengklasifikasikan sentimen.

Random Forest memiliki kemampuan untuk mengurangi *overfitting*, serta meningkatkan akurasi prediksi melalui kombinasi hasil dari banyak pohon. Random forest juga memiliki kemampuan yang baik untuk menangani *noise* dan variasi dalam data. Parameter seperti jumlah *tree* dan kedalaman *tree* dalam Random Forest memiliki pengaruh terhadap hasil prediksi. Jumlah *tree* yang lebih banyak dapat meningkatkan stabilitas dari model dan kedalaman *tree* yang optimal dapat membantu menangkap pola dalam data tanpa menyebabkan *overfitting*.

2. LANDASAN KEPUSTAKAAN

2.1. Analisis Sentimen

Analisis sentimen atau *opinion mining* adalah bidang ilmu komputasi yang berguna untuk melakukan analisis opini, sentimen, evaluasi, penilaian, sikap, dan emosi dari pernyataan seseorang yang terdapat dalam suatu teks (Liu, 2012). Dengan menggunakan analisis sentimen, maka dapat menghemat waktu dan tenaga ketika melakukan suatu penelitian atau analisis data yang melibatkan jumlah data yang besar.

2.2. Artificial Intelligence (AI)

Menurut KBBI, *Artificial Intelligence* merupakan program komputer yang dapat meniru kecerdasan manusia untuk melakukan hal-hal seperti mengambil keputusan, membuat dasar penalaran, dan karakteristik manusia lainnya. *Artificial Intelligence* juga dapat didefinisikan sebagai sebuah program komputer yang dilengkapi dengan algoritma yang memungkinkan untuk mempelajari data dan menggunakan pengetahuannya tersebut untuk melakukan pemikiran dan bertindak seperti yang dilakukan manusia (Rouhiainen, 2018).

2.3. X/Twitter

X/Twitter merupakan sebuah layanan *microblogging* yang penggunanya dapat membuat serta membaca pesan secara *real time* yang biasa disebut dengan *tweet*. Biasanya penggunaan X/Twitter akan semakin meningkat apabila sedang terjadi suatu peristiwa-peristiwa yang populer.

2.4. Text Preprocessing

Text preprocessing merupakan tahap yang penting dalam analisis data yang melibatkan seleksi dan pengolahan data teks untuk menjadikan data tersebut menjadi lebih terstruktur melalui beberapa tahap (Sinurat, et al., 2023). Tahap *text preprocessing* berguna untuk mempercepat suatu proses komputerisasi, karena diperlukan adanya seleksi pada data yang digunakan, sehingga struktur data tersebut menjadi lebih baik dan lebih siap untuk tahap selanjutnya. Pada umumnya dalam proses *text preprocessing* terdapat dari beberapa tahap, seperti *cleaning*, *case folding*, *tokenizing*, *filtering*, *stemming* (Audrey, et al., 2022).

2.5. Pembobotan Kata

Metode Term Frequency-Inverse Document Frequency (TF-IDF) dipakai dalam menghitung nilai relatif setiap kata dalam suatu data dokumen. *Term frequency* (TF) menunjukkan dalam suatu dokumen berapa frekuensi kemunculan suatu kata, sementara *document frequency* (DF) menunjukkan jumlah dari dokumen yang memuat kata tersebut (Barus, 2023). Berikut ini rumus TF pada Persamaan (1).

$$TF_{(t,d)} = \frac{\text{Jumlah kemunculan kata } t}{\text{Total kata dalam dokumen } d}$$

(1)

Berikut ini rumus IDF pada Persamaan (2).

$$IDF_{(t)} = \log \left(\frac{N}{DF_{(t)}} \right)$$

(2)

Berikut ini rumus TF-IDF pada Persamaan (3).

$$TF - IDF_{(t,d)} = TF_{(t,d)} \times IDF_{(t)}$$

(3)

2.6. Random Forest

Random Forest adalah metode yang dapat dimanfaatkan dalam regresi maupun klasifikasi. Random Forest dikembangkan dari klasifikasi dasar Decision Tree. Metode Random Forest melibatkan penggunaan banyak pohon keputusan yang menjadi dasar klasifikasi yang kemudian akan digabungkan (Afdhal, et al., 2022). Penggunaan pohon yang semakin banyak memiliki efek yang lebih baik pada akurasi yang dapat dicapai.

Pembuatan Decision Tree dapat diawali dengan melakukan perhitungan nilai *gini impurity* dan kemudian dilanjutkan dengan perhitungan *average gini impurity* yang dapat menentukan nilai *information gain* dan tingkat informatif dari sebuah node atribut (Basar, et al., 2023). Berikut ini rumus *gini impurity* pada Persamaan (4).

$$Gini\ Impurity = 1 - \sum_{i=1}^n (P_i)^2$$

(4)

Berikut ini rumus *average gini impurity* pada Persamaan (5).

$$Average\ Gini\ Impurity = \frac{n_{false}}{i} \times GI\ False + \frac{n_{true}}{i} \times GI\ True$$

(5)

Berikut ini rumus *information gain* pada Persamaan (6).

$$Information\ Gain =$$

$$Gini\ Impurity - Average\ Gini\ Impurity$$

(6)

2.7. Confusion Matrix

Confusion Matrix merupakan tahapan penting dalam *data mining* untuk menghitung akurasi dari suatu model (Han, et al., 2006). Confusion Matrix banyak digunakan dalam penelitian untuk melakukan evaluasi hasil dan pengukuran kemampuan dari metode yang digunakan. Berikut ini Confusion Matrix dengan dua kelas seperti pada Tabel 1.

Tabel 1. Confusion Matrix

Kelas	Prediksi Positif	Prediksi Negatif
Aktual Positif	<i>True Positive</i> (TP)	<i>False Negative</i> (FN)
Aktual Negatif	<i>False Positive</i> (FP)	<i>True Negative</i> (TN)

2.8. Fleiss Kappa

Fleiss Kappa adalah salah satu metrik statistik yang digunakan untuk melakukan pengukuran tingkat kesepakatan atau kecocokan antara tiga atau lebih *annotator* yang mengklasifikasikan objek ke dalam kategori yang sama atau berbeda (Fleiss, et al., 2003). Fleiss Kappa dapat digunakan untuk mengevaluasi sejauh mana para *annotator* melakukan klasifikasi secara konsisten. Nilai Fleiss Kappa dapat didapatkan dengan melakukan perhitungan pada *observed agreement* (Po), *expected agreement* (Pe), dan koefisien Fleiss Kappa. Nilai Pe, Po, dan koefisien Fleiss Kappa dapat diperoleh masing-masing dengan Persamaan (7), Persamaan (8), dan Persamaan (9).

$$Pe = \sum_{j=1}^k p_j^2$$

(7)

$$Po = \frac{1}{N \times n \times (n-1)} (\sum_{i=1}^N \sum_{j=1}^k n_{ij}^2 - N \times n)$$

(8)

$$k = \frac{Po - Pe}{1 - Pe}$$

(9)

2.9. Word Cloud

Word Cloud adalah teknik visualisasi data tekstual secara grafis berdasarkan frekuensi kemunculan kata yang dipresentasikan secara informatif dan menarik (Pradana, 2020). Semakin besar ukuran kata, maka dapat diartikan bahwa semakin sering kata tersebut muncul dalam sebuah dokumen. Penggunaan *Word Cloud* bertujuan untuk memberikan pemahaman cepat mengenai kata-kata yang dominan dalam data yang sedang dianalisis. *Word Cloud* dapat menjadi alat yang berguna untuk memperoleh wawasan yang lebih mendalam terhadap isi data yang akan yang dianalisis.

3. METODOLOGI

3.1. Tipe Penelitian

Tipe penelitian yang diterapkan pada analisis sentimen pada penelitian ini adalah tipe non-implementatif dengan pendekatan analitik. Tipe penelitian non-implementatif mengacu pada fokus penelitian yang terletak pada investigasi terhadap sebuah peristiwa maupun fenomena untuk menghasilkan sebuah analisis ilmiah.

3.2. Strategi Penelitian

Dalam penelitian ini akan menghasilkan analisis sentimen dari opini masyarakat terkait dengan dampak perkembangan *Artificial Intelligence*. Penelitian ini menggunakan data *tweet* dan diklasifikasikan ke dalam dua kelas sentimen, yakni positif serta negatif. Pengklasifikasian sentimen dilaksanakan dengan metode Random Forest, sementara metode Confusion Matrix digunakan untuk evaluasi hasil klasifikasi.

3.3. Subjek Penelitian

Penelitian ini berfokus pada analisis sentimen terhadap *tweet* yang dibuat oleh pengguna X/Twitter. Dengan melibatkan pengguna X/Twitter sebagai subjek penelitian, diharapkan dapat memberikan wawasan mengenai pendapat, perasaan, dan perspektif dari pengguna X/Twitter terkait dengan dampak perkembangan *Artificial Intelligence*.

3.4. Tahap Penelitian

Pada penelitian ini dilakukan serangkaian tahapan penelitian, seperti pengumpulan data, pelabelan data, *text preprocessing*, pembobotan kata, klasifikasi sentimen, analisis dan

pengujian, serta penarikan kesimpulan dan saran. Tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

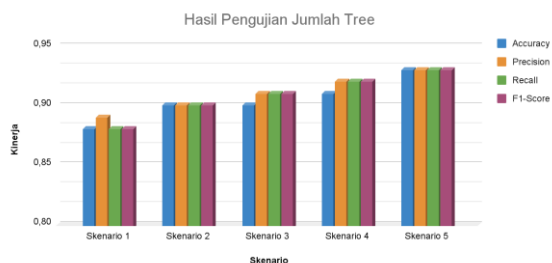
4. ANALISIS DAN PENGUJIAN

4.1. Fleiss Kappa

Kesepakatan hasil pelabelan dilakukan evaluasi menggunakan Fleiss Kappa, yang merupakan suatu metode untuk mengukur tingkat kesepakatan antara tiga atau lebih *annotator*. Hasil Fleiss Kappa menunjukkan tingkat reliabilitas yang hampir sempurna dengan nilai sebesar 0,98. Nilai tersebut menunjukkan tingkat kesepakatan yang hampir sempurna antara tiga *annotator* dalam proses pelabelan. Nilai tersebut juga menunjukkan bahwa proses pelabelan data dilakukan secara konsisten dan akurat oleh para *annotator* yang terlibat.

4.2. Pengujian dan Analisis Jumlah Tree

Pengujian menggunakan jumlah *tree* sebagai parameter untuk mengetahui bagaimana parameter yang dipakai berpengaruh pada hasil klasifikasi. Terdapat 5 skenario jumlah *tree* yang setiap skenarionya dijalankan sejumlah 5 kali. Parameter kedalaman *tree* yang digunakan setiap skenario bernilai tetap yaitu 15. Visualisasi hasil pengujian jumlah *tree* dapat dilihat pada Gambar 2.



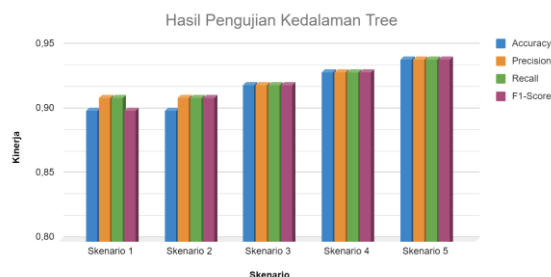
Gambar 2. Visualisasi Hasil Pengujian Jumlah Tree

Berdasarkan Gambar 2 dapat dilihat terdapat peningkatan rata-rata nilai *accuracy*, *precision*, *recall*, serta *f1-score* pada skenario ke-1 hingga skenario ke-5. Nilai terbaik pada pengujian jumlah *tree* tersebut terdapat pada skenario ke-5

dengan jumlah *tree* yaitu 60 serta rata-rata nilai *accuracy* 0,93, *precision* 0,93, *recall* 0,93, dan *f1-score* 0,93. Berdasarkan hasil pengujian yang didapat, parameter jumlah *tree* berpengaruh terhadap hasil prediksi, semakin besar jumlah *tree* yang digunakan, maka akan terjadi peningkatan kesempatan untuk membuat prediksi yang lebih tepat dan akurat.

4.3. Pengujian dan Analisis Kedalaman Tree

Pengujian menggunakan kedalaman *tree* sebagai parameter untuk mengetahui bagaimana kedalaman *tree* yang digunakan berpengaruh terhadap hasil prediksi. Terdapat 5 skenario kedalaman *tree* yang masing-masing dijalankan sejumlah 5 kali. Parameter jumlah *tree* yang dipakai setiap skenario bernilai tetap yaitu 60. Visualisasi hasil pengujian kedalaman *tree* dapat dilihat pada Gambar 3.



Gambar 3. Visualisasi Hasil Pengujian Kedalaman Tree

Berdasarkan Gambar 3 dapat dilihat terdapat peningkatan rata-rata nilai *accuracy*, *precision*, *recall*, serta *f1-score* pada skenario ke-1 hingga skenario ke-5. Nilai terbaik pada pengujian kedalaman *tree* tersebut terdapat pada skenario ke-5 dengan kedalaman *tree* yaitu 55 serta rata-rata nilai *accuracy* 0,94, *precision* 0,94, *recall* 0,94, serta *f1-score* 0,94. Berdasarkan pengujian yang dilakukan, parameter kedalaman *tree* berpengaruh terhadap hasil prediksi, semakin besar batas maksimal kedalaman *tree*, maka *tree* yang dibentuk akan lebih kompleks dan dapat menangkap lebih banyak informasi dari data yang dapat meningkatkan kesempatan untuk membuat prediksi yang lebih tepat dan akurat.

4.4. False Positive dan Negative

Dalam analisis sentimen kesalahan dapat terjadi dalam dua bentuk, yaitu *false positive* (FP) dan *false negative* (FN). *False positive* terjadi ketika model salah mengklasifikasikan sentimen negatif sebagai positif, sedangkan *false*

negative terjadi ketika model salah mengklasifikasikan sentimen positif sebagai negatif. Kesalahan *false positive* (FP) dan *false negative* (FN) dalam analisis sentimen dapat terjadi karena kurangnya pemahaman konteks yang sebenarnya terkandung dalam *tweet*, sehingga Model tidak mampu memahami konteks negatif yang sebenarnya dari *tweet* tersebut.

Kesalahan *false positive* (FP) dan *false negative* (FN) juga dapat dipengaruhi oleh kurangnya variasi opini yang mencakup rentang yang lebih luas dalam data pelatihan. Kurangnya variasi opini dalam data pelatihan dapat menyebabkan model tidak dapat mempelajari dengan baik konteks tertentu dalam *tweet*, sehingga menghasilkan kesalahan dalam klasifikasi. Data pelatihan yang hanya mencakup *tweet* yang dominan positif atau negatif membuat model mungkin akan mengklasifikasikan *tweet* yang seharusnya termasuk dalam kelas sentimen yang berbeda.

4.5. Visualisasi Word Cloud

Word Cloud digunakan untuk melakukan visualisasi data teks dan menunjukkan kata-kata yang muncul dalam jumlah tinggi pada suatu dokumen. Ukuran sebuah kata-kata pada *Word Cloud* menggambarkan seberapa sering kata tersebut muncul pada sebuah dokumen, sehingga memungkinkan untuk melakukan identifikasi kata-kata yang paling dominan dengan cepat. Semakin besar ukuran kata, maka kata tersebut memiliki frekuensi kemunculan yang lebih tinggi. Hasil *word cloud* sentimen positif dapat dilihat pada Gambar 4.



Gambar 4. Hasil *Word Cloud* Sentimen Positif

Berdasarkan Gambar 4 dapat dilihat terdapat kata-kata berukuran besar seperti “membantu”, “meningkatkan”, “efisiensi”, “manusia”, “perkembangan”, “canggih”, dan “mempermudah” menggambarkan sentimen positif terkait dampak perkembangan *Artificial Intelligence* (AI). Kata-kata tersebut

menunjukkan bahwa banyak pengguna yang melihat AI sebagai alat yang dapat meningkatkan efisiensi dalam berbagai aspek kehidupan. kata-kata seperti "perkembangan" dan "canggih" menandakan bahwa AI dianggap sebagai teknologi yang mampu mendorong kemajuan dan inovasi dalam berbagai bidang. Hasil *Word Cloud* ini menggambarkan bahwa AI dipandang sebagai sebuah kemajuan teknologi yang memiliki potensi besar untuk meningkatkan kualitas hidup dan membantu memecahkan tantangan-tantangan yang dihadapi oleh manusia.

Perlu juga diperhatikan hasil *Word Cloud* sentimen negatif untuk melengkapi gambaran terkait dampak perkembangan *Artificial Intelligence* (AI). Hasil *word cloud* sentimen negatif dapat dilihat pada Gambar 5.



Gambar 5. Hasil *Word Cloud* Sentimen Negatif

Berdasarkan Gambar 5 dapat dilihat terdapat kata-kata berukuran besar seperti “ngeri”, “serem”, “bahaya”, “foto”, “takut”, dan “hoax” menggambarkan sentimen negatif terkait dampak perkembangan *Artificial Intelligence* (AI). Kata-kata tersebut menunjukkan bahwa terdapat pengguna yang merasa cemas atau takut terhadap potensi bahaya yang mungkin dapat ditimbulkan akibat penggunaan AI. Adanya kekhawatiran terhadap privasi dan penyebaran informasi palsu juga menjadi faktor yang memengaruhi sentimen negatif tersebut. Hasil *Word Cloud* ini menunjukkan pentingnya memperhatikan aspek etika, keamanan, dan dampak sosial dari penggunaan teknologi AI untuk mengurangi risiko dan memastikan bahwa teknologi ini digunakan secara bertanggung jawab.

5. PENUTUP

5.1. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, maka dapat diperoleh beberapa kesimpulan yaitu:

1. Random Forest bisa menjadi pilihan metode guna mengklasifikasi data pada X/Twitter. Penggunaan metode Random Forest dapat dilakukan dalam beberapa tahapan, yaitu pengumpulan data dengan *web scraping* menggunakan *library* Tweet Harvest, *text preprocessing* data, pembobotan kata dengan metode *Term Frequency-Inverse Document Frequency* dan memanfaatkan *library* Scikit-Learn, serta klasifikasi menggunakan Random Forest dan memanfaatkan *library* Scikit-Learn.
2. Evaluasi klasifikasi Random Forest menggunakan metode confusion matrix dengan parameter yaitu jumlah *tree* 60 dan kedalaman *tree* 55 memperoleh *accuracy* 0,94, *precision* 0,94, *recall* 0,94, dan *f1-score* 0,94. Berdasarkan hasil yang ada menunjukkan model klasifikasi memiliki kinerja yang baik. Kesalahan yang terjadi pada *false positive* (FP) dan *false negative* (FN) dapat disebabkan karena kurangnya pemahaman konteks yang sebenarnya terkandung dalam *tweet*, dan juga dapat disebabkan karena kurangnya variasi opini yang mencakup rentang yang lebih luas dalam data pelatihan.
3. Hasil *Word Cloud* terkait dampak perkembangan *Artificial Intelligence* (AI) menunjukkan persepsi yang beragam. Hasil kata yang dominan pada sentimen positif menunjukkan bahwa AI dapat membantu serta memudahkan manusia dalam berbagai aspek, sedangkan pada sentimen negatif menunjukkan kekhawatiran akan bahaya dan risiko terhadap penggunaan AI. Perbedaan ini mencerminkan kompleksitas dan tantangan dalam membuat kebijakan dan mengelola teknologi AI serta pentingnya memahami dan mengatasi berbagai risiko yang terkait.

5.2. Saran

Berdasarkan penelitian yang telah dilakukan, maka saran yang dapat diberikan untuk penelitian selanjutnya yaitu:

1. Penelitian berikutnya bisa menggunakan metode klasifikasi lain untuk menjadi bahan perbandingan terhadap metode digunakan pada penelitian ini yaitu Random Forest.
2. Penelitian selanjutnya dapat menggunakan metode Random Forest dan menggunakan data dengan jumlah lebih banyak.

3. Penelitian selanjutnya dapat mengumpulkan data terkait topik yang sama menggunakan media sosial selain X/Twitter, seperti Instagram, Facebook, Youtube, Tiktok, maupun media sosial lain yang terkait.

6. DAFTAR PUSTAKA

Aditya, B. R. (2015). Penggunaan Web Crawler Untuk Menghimpun Tweets dengan Metode Pre-Processing Text Mining. *Jurnal Infotel*, 7(2), 93-100.

Afdhal, I., Kurniawan, R., Iskandar, I., Salambue, R., Budianita, E., & Syafria, F. (2022). Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia. *Jurnal Nasional Komputasi dan Teknologi Informasi*, 5(1), 122-130.

Anugrah, Z. (2023, Juli 26). Dampak Positif dan Negatif dari Kecerdasan Buatan (AI) dalam Kehidupan. *OSC Medcom*, [online] Tersedia di: <<https://osc.medcom.id/community/dampak-positif-dan-negatif-dari-kecerdasan-buatan-ai-dalam-kehidupan-5986>> [Diakses 23 Agustus 2023].

Audrey, O., Ratnawati, D. E., & Arwani, I. (2022). Analisis Sentimen Pengguna Twitter Terhadap Opini Non Fungible Token di Indonesia Menggunakan Algoritma Random Forest Classifier. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 6(12), 5889-5897.

Barus, T. P. (2023). Analisis Sentimen pada PLN Mobile Menggunakan Metode Decision Tree. *S1. Universitas Medan Area*.

Basar, T. F., Ratnawati, D. E., & Arwani, I. (2022). Analisis Sentimen Pengguna Twitter terhadap Pembayaran Cashless menggunakan Shopeepay dengan Algoritma Random Forest. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 6(3), 1426-1433.

Fitri, E. (2020). Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector

- Machine. *Jurnal Transformatika*, 18(1), 71-80.
- Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions* (3rd ed.). Hoboken, NJ: Wiley.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques* (3rd ed), Elsevier Inc. <https://doi.org/10.1016/C2009-0-61819-5>
- Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers.
- Mahardhika, Y. S., & Zuliarso, E. (2018). Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritma Naives Bayes Classifier. *Prosiding SINTAK*.
- Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012). How many trees in a random forest?. *Machine Learning and Data Mining in Pattern Recognition* (pp. 154-168). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-31537-4_13.
- Pakpahan, R. (2021). Analisa Pengaruh Implementasi Artificial Intelligence Dalam Kehidupan Manusia. *JISICOM (Journal of Information System, Informatics and Computing)*, 5(2), 506-513.
- Pradana, M. G. (2020). Penggunaan Fitur Wordcloud dan Document Term Matrix dalam Text Mining. *Jurnal Ilmiah Informatika*, 8(01), 38-43.
- Rouhiainen, L. (2018). *Artificial Intelligence: 101 things you must know today about our future*. CreateSpace Independent Publishing Platform.
- Sinurat, D. M. Y., Ratnawati, D. E., & Brata, D. W. (2023). Analisis Sentimen Terhadap Kenaikan Cukai Rokok pada Media Sosial Twitter menggunakan Algoritma Naive Bayes Classifier. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 7(1), 17-25.
- Sudianto., Wahyuningtias, P., Utami, H. W., Raihan, U. A., Hanifah, H. N., & Adanson, Y. N. (2022). Perbandingan Metode Random Forest dan Support Vector Machine pada Analisis Sentimen Twitter (Studi Kasus: Kaburnya Selebgram Rachel Venny dari Karantina). *Jurnal Teknik Informatika (JUTIF)*, 3(1), 141-145.
- Yanti, C. P., Agustini, N. W. E., Ginantra, N. L. W. S. R., & Wulandari, D. A. P. (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), 756-765. <https://doi.org/10.30865/mib.v7i2.5900>.