

ANALISIS SENTIMEN PERAN ARTIFICIAL INTELLIGENCE TERHADAP KREATIVITAS DAN EFEKTIVITAS MAHASISWA DALAM PENYELESAIAN TUGAS AKHIR MENGGUNAKAN DECISION TREE BERBASIS SMOTE

Indi Alya Putri, Andri Firmansyah, Asep Suprianto

S1 Teknik Informatika, Universitas Pelita Bangsa

Jl. Inspeksi Kalimalang No.9, Cibatu, Cikarang Selatan, Kab Bekasi, Jawa Barat

Indyputri44@gmail.com

ABSTRAK

Dalam penelitian mengenai penggunaan *Artificial Intelligence* (AI) untuk mendukung penyelesaian tugas perkuliahan hingga membantu proses penyelesaian tugas akhir. Akan tetapi dengan banyak kemudahan yang diberikan oleh AI mahasiswa terkadang melupakan aspek terpenting dalam karya ilmiah yaitu bersifat orisinal, berbagai opini yang muncul dari mahasiswa mengenai penggunaan AI dapat membantu proses penelitian menjadi lebih cepat akan tetapi meningkatkan resiko *plagiarism*. Oleh karena itu, penelitian ini bertujuan untuk mengidentifikasi berbagai opini mahasiswa mengenai penggunaan AI, baik dari sisi positif maupun negatif, serta melihat persentase mayoritas. Melalui analisis dampak yang ditimbulkan, penelitian ini diharapkan dapat memberikan wawasan yang berguna bagi Universitas dan pembuat kebijakan dalam menetapkan batasan dan etika-etika penggunaan AI yang sesuai. Proses analisis sentimen menggunakan algoritma *Decision Tree* dan untuk menyempurnakan model tersebut dilengkapi dengan teknik *Synthetic Minority Over-sampling Technique SMOTE* untuk mengatasi data yang tidak seimbang. Dalam analisis sentimen, kata yang paling sering muncul adalah positif, AI, lebih, membantu. Berdasarkan analisis dengan *Decision Tree* berbasis *SMOTE*, diperoleh akurasi sebesar 88%, dengan nilai presisi, recall, dan F1-Score mencapai 91% dengan hasil analisis sentimen menunjukkan bahwa 69,7% mahasiswa memberikan sentimen positif terhadap penggunaan AI, sementara 30,3% negatif. Temuan ini mengindikasikan bahwa mayoritas mahasiswa merasa terbantu oleh AI dalam penyelesaian tugas akhir.

Kata kunci : *Artificial Intelligence, Decision Tree, Sentimen Analisis, Skripsi, SMOTE*

1. PENDAHULUAN

Teknologi Informasi, yang begitu pesat di era serba digitalisasi ini menjadi faktor yang sangat berpengaruh, dalam membawa perubahan besar bagi dunia pendidikan. Kemajuan teknologi, khususnya dalam bidang *Artificial Intelligence* (AI), telah menghadirkan sebuah solusi inovatif yang mempermudah mahasiswa untuk memahami pembelajaran dan memberi kemudahan.

Di dunia pendidikan, AI memainkan peran penting dalam membantu mahasiswa mengerjakan tugas, bahkan hingga menyelesaikan tugas akhir yaitu skripsi. Teknologi ini mampu menyediakan akses cepat untuk mendapatkan informasi, membantu memilih judul dan menambah ide baru yang dapat membantu penelitian berjalan, mengolah data dalam jumlah besar. Namun, ditengah kemudahan yang dibantu oleh teknologi sekarang ini, muncul banyak opini publik mengenai etika dan penggunaan AI, terutama dalam membantu pembuatan karya ilmiah dan skripsi mahasiswa, dimana banyak masyarakat berpendapat bahwa penggunaan AI dalam pengerjaan skripsi dapat menyebabkan kemalasan pada mahasiswa, mengurangi berpikir kritis dan bahkan dianggap bentuk dari *plagiarism* karena terlalu bergantung pada teknologi tersebut.

Penelitian ini penting dilakukan karena, meskipun banyak yang mengakui kemampuan AI dalam mendukung efisiensi di bidang akademik, masih

terdapat keraguan mengenai pengaruhnya terhadap kualitas hasil karya ilmiah yang dihasilkan dan integritas akademik mahasiswa. Oleh karena itu, dengan analisa sentimen mahasiswa terhadap pemanfaatan AI, penelitian ini bertujuan untuk menggali lebih dalam pandangan mereka. Temuan dari penelitian ini akan memberikan landasan yang berguna bagi Universitas dalam mengambil kebijakan yang tepat mengenai etika penerapan teknologi AI dalam dunia pendidikan.

Selain itu, penelitian ini juga penting untuk mengukur seberapa efektif dampak AI dalam membantu mahasiswa dalam menyelesaikan tugas akhir, seperti mendukung proses penelitian, penulisan serta mengembangkan ide-ide baru.

Penelitian ini mengklasifikasikan berbagai opini mahasiswa terkait penggunaan AI dalam membantu pengerjaan skripsi. Menggunakan klasifikasi algoritma *decision tree* (pohon keputusan). Klasifikasi merupakan teknik data mining untuk memetakan data atau label kelas yang telah ditentukan.

Analisis sentimen salah satu cabang dari text mining dengan mengklasifikasi text untuk memahami opini atau pendapat yang terkandung di dalamnya [1].

Analisis sentimen bertujuan mengidentifikasi berupa emosi, opini dalam text, seperti pendapat atau ulasan terkait suatu topik, dalam menganalisis sentimen dapat menggunakan *NLP* sebagai alat pemrosesan bahasa alami [2].

Dalam penelitian ini pentingnya untuk menganalisis sentimen dari mahasiswa guna untuk memahami persepsi mahasiswa terhadap penggunaan tools AI yang populer di kalangan mahasiswa dalam mempermudah proses pembelajaran dan penyelesaian tugas akhir. Output yang diperoleh dalam analisis ini beragam ada opini positif maupun negatif, sehingga dapat mengoptimalkan penggunaan AI sesuai dengan preferensi mahasiswa serta memahami etika penggunaan AI.

Studi terdahulu, penelitian Fatkhudin [3], menunjukkan pentingnya menganalisis sentimen untuk memahami persepsi masyarakat terkait penggunaan AI. Serta memanfaatkan algoritma *decision tree* C4.5 guna mengklasifikasi menjadi kategori positif maupun negatif. Hasil klasifikasi opini masyarakat dari social media x diperoleh 84,4% berisi sentimen negatif dan 15,6% sentimen positif, dengan akurasi model *decision tree* yang cukup akurat mencapai 66%.

Selain itu penggunaan model *decision tree* ini dapat dikatakan efektif terbukti dengan meningkatkan nilai akurasi dalam analisis sentimen [4].

Teknik yang digunakan untuk menganalisis seperti *decision tree* berbasis SMOTE kombinasi yang menarik mengingat penggunaan SMOTE yang mampu mengatasi terjadinya bias pada dataset atau dapat menangani data yang tidak seimbang [5].

Pada penelitian yang berjudul analisis peran *Artificial Intelligence* (AI), khususnya *ChatGPT* dalam perkuliahan di kalangan mahasiswa PGSD UPI dengan pendekatan kualitatif dan desain studi kasus. Melalui kuesioner yang disebar ke 30 mahasiswa, penelitian ini menemukan bahwa Sebagian besar responden familiar dengan *ChatGPT* dan menggunakannya untuk membantu tugas serta dapat menghemat waktu. Namun, terdapat kekhawatiran bahwa kemudahan tersebut dapat mengurangi kemampuan berpikir kritis mahasiswa [6].

2. TINJAUAN PUSTAKA

Pada bagian ini memuat literatur ilmiah yang relevan dengan penelitian yang dilakukan saat ini

2.1. Artificial Intelligence

Artificial Intelligence atau kecerdasan buatan adalah sebuah sistem komputer yang berfokus pada kecerdasan manusia, perangkat yang akan memahami lingkungan sekitar dan melakukan tugas-tugas yang membutuhkan proses berpikir manusia dan mendesain mesin agar dapat meniru perilaku manusia [7].

2.2. Analisis Sentimen

Analisis sentimen merupakan salah satu cabang dari text mining yang berfokus pada sebuah proses dari klasifikasi pada dokumen text untuk memahami opini atau pendapat yang terkandung di dalamnya [8].

Dengan melakukan analisis sentimen, kita dapat mengevaluasi issue atau permasalahan, dan menentukan pendapat tentang masalah, apakah

permasalahan tersebut mempunyai kecenderungan positif, negatif, atau netral.

2.3. Decision Tree

Decision tree merupakan Teknik analisis data untuk menentukan hasil dengan membentuk model klasifikasi yang berbentuk struktur pohon. Data akan dibagi menjadi bagian-bagian lebih kecil [9].

Mengelaborasi cabang yang membawa noise [10]. Serta membantu mengarahkan berbagai kemungkinan yang menggambarkan pertanyaan atau pengujian terhadap atribut tersebut.

2.4. SMOTE

Synthetic Minority Over-sampling Technique (SMOTE) metode yang sering digunakan untuk mengatasi ketidakseimbangan jumlah data pada kategori dengan cara menghasilkan data sintesis pada kelas yang minor, sehingga distribusi data menjadi lebih seimbang antara kategori kelas minor dan mayor [11].

Dengan menambahkan variasi data, mengurangi dominasi kelas mayoritas, sehingga mengurangi resiko duplikasi berlebihan yang dapat menyebabkan overfitting.

2.5. Kreativitas

Kreativitas adalah eksplorasi ide yang meningkatkan pengembangan dan membantu memecahkan masalah serta menemukan solusi kreatif. Kreativitas mampu menciptakan inovasi baru yang *out-of-the-box*. Dalam dunia pendidikan, kreativitas memfasilitasi proses pembelajaran yang aktif dan seru, sehingga dapat menciptakan kurikulum inovatif [12].

2.6. Efektivitas

Penggunaan teknologi ini mampu meningkatkan efektivitas dan hasil belajar siswa secara signifikan, mampu meningkatkan kemampuan *brainstorming*, kalimat kalimat yang sebelumnya tidak terpikirkan serta kemampuan dalam berpikir kritis. Selain itu juga untuk menghemat waktu pembelajaran untuk bisa memahami lebih cepat jika memanfaatkan teknologi di karenakan mendapat berbagai jenis informasi dari berbagai sumber [13].

2.7. Google Collab

Google Colab merupakan salah satu *Integrated Development Environment* (IDE) berbasis web yang dirancang untuk memprogram menggunakan bahasa *Python* yang pemrosesan berbasis *cloud* yang sudah dilengkapi dengan sebagian besar pustaka *Python* yang sering digunakan, seperti *NumPy*, *pandas*, dan *TensorFlow*, sehingga memudahkan pengguna untuk langsung memulai analisis data atau pengembangan model pembelajaran mesin tanpa instalasi [14].

2.8. Teks Mining

Teks mining merupakan metode untuk menggali inti dari sebuah dokumen teks yang memiliki struktur

tidak teratur, sehingga kemudian akan dianalisis untuk menghasilkan informasi yang bermanfaat [15]. Proses data mining bisa berasal dari banyaknya sumber, termasuk teks dari hasil kuesioner yang akan dianalisis pola, topik, atau sentimen yang terkandung dalam jawaban responden

2.9. TF-IDF

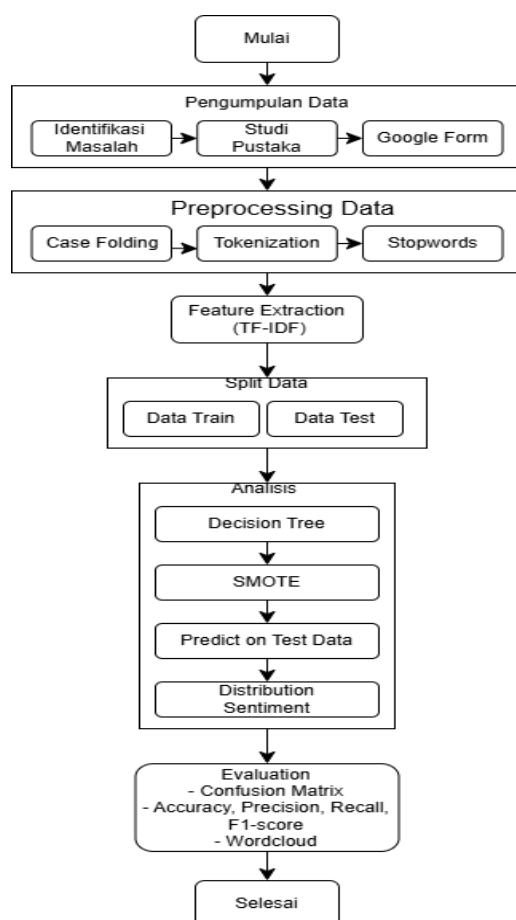
TF-IDF adalah teknik pembobotan kata dengan menghitung nilai *Term Frequency* kemunculan kata-kata populer atau yang sering muncul dalam teks dokumen [16].

Serta mengubah teks menjadi data dalam bentuk numerik agar lebih mudah diproses dan dianalisis. Dengan metode ini, TF-IDF membantu model untuk memfokuskan kata-kata yang memiliki relevansi tinggi, sehingga mampu meningkatkan akurasi dalam klasifikasi sentimen.

2.10. Evaluasi Model

Proses evaluasi menggunakan *cross-validation* dan perhitungan manual *confusion matrix* mencakup pengukuran *akurasi*, *presisi*, *recall*, dan *F1-Score* pada data testing. Proses evaluasi ini krusial untuk memastikan bahwa model memberikan hasil yang tepat dan relevan.

3. METODE PENELITIAN



Gambar 1. Alur penelitian

Penelitian ini, dimulai dengan pengumpulan data melalui data kuesioner kepada mahasiswa semester akhir yang sedang menyelesaikan tugas akhir guna untuk melihat persepsi mereka terkait penggunaan AI populer dikalangan mahasiswa. Proses penelitian disajikan pada gambar 1.

3.1. Metode Pengumpulan Data

Dataset dalam penelitian ini diperoleh melalui penyebaran kuesioner menggunakan Google Form kepada 328 mahasiswa aktif Teknik informatika Angkatan 2021 di Universitas Pelita Bangsa. Dengan total responden yang di dapat sebanyak 180 orang.

3.2. Pengambilan Sampel

Dalam penelitian ini untuk mengambil sampel menggunakan metode *simple random sampling* digunakan untuk menentukan populasi yang akan diambil untuk dianalisis pada tahap selanjutnya, dimana setiap populasi memiliki kesempatan terpilih menjadi sample [17].

Maka untuk perhitungan sample dapat menggunakan rumus slovin sebagai berikut

$$n = \frac{N}{1 + N(e)^2} \quad (1)$$

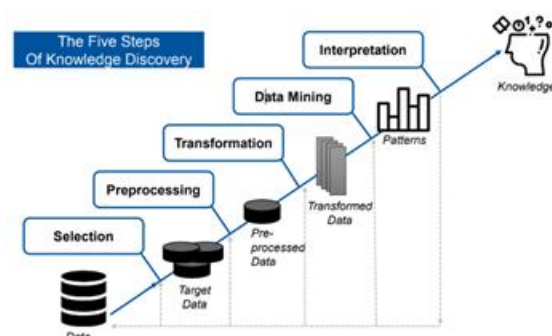
Dalam persamaan (1) ini, n menunjukkan jumlah sample yang akan digunakan, sedangkan N itu jumlah dari keseluruhan populasi, lalu e merupakan batas atau nilai margin of error masih bisa diterima atau ditoleransi [17].

3.3. Metode KDD

Tahapan penelitian yang disusun secara sistematis guna untuk membuat proses penelitian berjalan dengan lancar dan terstruktur. Dalam sebuah penelitian digunakan metodologi *Knowledge Discovery in Database* (KDD) sebagai panduan dalam setiap tahapnya [18].

Berfokus pada pengumpulan data, mengidentifikasi pola, dan proses pengujian. Serta memanfaatkan aplikasi Google Colab untuk mengolah dan mengevaluasi data. Dalam penelitian ini, diperlukan Langkah-langkah pengolahan data mengikuti metode *Knowledge Discovery in Database* (KDD), dengan tujuan untuk menghasilkan informasi yang terstruktur dan relevan dengan penelitian [19].

Visualisasi dari proses keseluruhan tahapan metodologi penelitian dapat dilihat pada Gambar 2.



Gambar 2. Metode knowledge discovery in database

3.4. Data Selection

Pada tahap *selection*. Dimulai dengan menentukan populasi penelitian atau proses pengumpulan data untuk dijadikan populasi dan sampel penelitian, setelah itu data dikumpulkan dengan penyebaran kuesioner kepada mahasiswa yang menjadi target reponden. Kuesioner pada penelitian ini dirancang mencakup dua jenis pertanyaan utama dengan skala likert. Jenis pertanyaan skala likert yang terdapat pada kuesioner penelitian, dapat dilihat pada Gambar 3.

Gambar 3. Skala likert

Pada *research question* juga diselipkan pertanyaan terbuka guna untuk memberikan ruang bagi responden menyampaikan opini mendalam terkait Penggunaan AI, serta pada pertanyaan ini juga responden dapat menentukan arah sentimen tersebut kategori sentimen positif maupun negatif.

Gambar 4. Jenis pertanyaan terbuka

3.5. Data Preprocessing

Preprocessing data merupakan tahap yang paling penting dalam pengolahan data untuk mengoptimalkan hasil analisis sentimen [20].

Proses ini penting untuk memastikan kebersihan data, tidak ada data duplikat, missing value, untuk menjadikannya data yang rapi dan terstruktur agar mendapatkan informasi yang akurat, serta lebih mudah dilakukan analisis data [9].

3.6. Data Transformation

Dalam proses analisis sentimen terhadap opini mahasiswa tentang penggunaan AI menggunakan algoritma *decision tree*, pada tahap transformation ini akan mengubah data yang telah diproses menjadi format numerik sehingga siap untuk dianalisis, dengan menggunakan *feature engineering* serta *vectorization*, di mana TF-IDF digunakan untuk menentukan kata-kata penting berdasarkan frekuensi kemunculannya.

3.7. Data Mining

Dalam menganalisis sentimen terhadap opini mahasiswa dalam penggunaan AI menggunakan

decision tree, tahap data mining dilakukan dengan mengumpulkan opini, serta melakukan Langkah *pre-processing* (*case folding*, *tokenization*, *stopwords*) serta pelabelan data, kemudian menerapkan algoritma decision tree untuk melatih model. Proses ini membantu memperoleh wawasan yang lebih mendalam mengenai pandangan mahasiswa terhadap suatu topik penelitian.

3.8. Evaluation

Pada tahap evaluasi ini meliputi data pengujian dan metrik seperti akurasi, presisi, *recall*, dan F1-score dengan bantuan *Confusion Matrix* untuk mengukur seberapa baik model dalam memprediksi sentimen, serta melihat seberapa tinggi nilai akurasi model decision tree dalam memahami opini mahasiswa.

4. HASIL DAN PEMBAHASAN

Penelitian ini memanfaatkan bahasa pemrograman Python dengan menggunakan *Google Collab* serta menggunakan algoritma *Decision Tree* dan *SMOTE* untuk membangun model analisis sentimen terhadap peran *Artificial Intelligence* dalam membantu mahasiswa dalam menyelesaikan tugas akhir. Dengan data yang diperoleh dari hasil penyebaran kuesioner yang peneliti lakukan.

4.1. Dataset

Dataset pada penelitian ini diperoleh dengan melakukan survei kepada responden, menggunakan *Google Formulir*, yang kemudian disebarluaskan melalui Grup Whats'app Teknik Informatika Angkatan 2021 Universitas Pelita Bangsa.

	D	E	F	G	H	I	J
1	Program s	Tingkat Pe	Semester s	AI_genera	pertanyaa	pertanyaa	pertan
2	Teknik Inf	Sering		7 ChatGPT	((	5	5
3	Teknik Inf	Sering		7 ChatGPT	((	5	5
4	Teknik Inf	Sering		7 Perplexity		5	5
5	Teknik Inf	Sering		7 ChatGPT	((	5	5
6	Teknik Inf	Sering		7 Claude		5	5
7	Teknik Inf	Kadang-ka		7 ChatGPT	((	3	3

Gambar 5. Dataset

Data yang dikumpulkan mencakup beberapa informasi penting seperti Prodi, Semester, Tingkat penggunaan *tools* AI, *tools* AI yang digunakan dan beberapa pertanyaan penelitian dengan skala likert serta pertanyaan terkait opini mereka terhadap penggunaan AI supaya lebih detail persepsi dari mahasiswa. Dengan data ini, diharapkan dapat di analisis dan memperoleh wawasan yang lebih jelas mengenai opini mahasiswa terhadap penggunaan AI yang semakin marak digunakan oleh mahasiswa untuk menyelesaikan tugas akhir yaitu skripsi. Serta memahami etika etika penggunaan AI dalam penulisan skripsi yang tetap mempertahankan orisinalitas.

4.2. Sampel Penelitian

Pada tahap pengambilan sampel dapat menggunakan rumus taro yamane, tahap ini diperlukan karena data yang diperoleh tidak semua dari jumlah mahasiswa aktif berpartisipasi dalam pengisian kuesioner Pada penelitian ini. Berdasarkan data, jumlah mahasiswa aktif yang menjadi populasi sebanyak 328 orang. Sedangkan, data yang diperoleh hanya 150 orang. Digunakan rumus slovin dengan perhitungan sebagai berikut :

$$n = \frac{N}{1 + N(e)^2}$$

$$n = \frac{328}{1 + 328(0.01)^2}$$

$$n = \frac{328}{4,28} = 76,6$$

$$n = 76$$

Berdasarkan perhitungan dengan rumus Taro Yamane, sampel penelitian ditetapkan tingkat presisi nilai error atau batas toleransi sebesar 10%, dengan populasi 328 orang. Setelah dilakukan perhitungan dapat ditarik kesimpulan untuk jumlah sampel yang diperlukan adalah 76 responden.

4.3. Preprocessing Data

Preprocessing merupakan tahap awal pada proses pengolahan data teks, hasil teks crawling sebelum diolah lebih lanjut hingga saat proses pelabelan data. Terdapat beberapa proses pada tahap ini yang terdiri dari case folding, cleaning, text duplikat, tokenizing, stopwords dan labeling. Berikut adalah alur dari data preprocessing.

4.4. Case Folding

Adalah proses untuk mengubah semua huruf kapital yang ada di dalam dataset menjadi huruf kecil. Tujuannya yaitu untuk data text yang di proses semuanya dalam bentuk huruf yang sama . berikut merupakan contoh hasil dari *case folding*.

Tabel 1. Case folding

Data Mentah	Data Setelah Case Folding
Lebih Positif Menurutku Karena Bisa Mengefisiensi Waktu Pengerjaan, Tapi Ya Bisa Jadi Negatif Karena Terlalu Bergantung	lebih positif menurutku karena bisa mengefisiensi waktu pengerjaan, tapi ya bisa jadi negatif karena terlalu bergantung

4.5. Tokenization

Merupakan proses untuk memecahkan kata yang sebelumnya berbentuk kalimat menjadi bentuk kata perkata. Berikut merupakan contoh hasil dari *tokenizing*.

Tabel 2. Tokenization

Data Mentah	Data Setelah Tokenization
lebih positif menurutku karena bisa mengefisiensi waktu pengerjaan, tapi ya bisa jadi negatif karena terlalu bergantung	'lebih', 'positif', 'menurutku', 'karena', 'bisa', 'mengefisiensi', 'waktu', 'pengerjaan', 'tapi', 'ya', 'bisa', 'jadi', 'negatif', 'karena', 'terlalu', 'bergantung'.

4.6. Stopwords

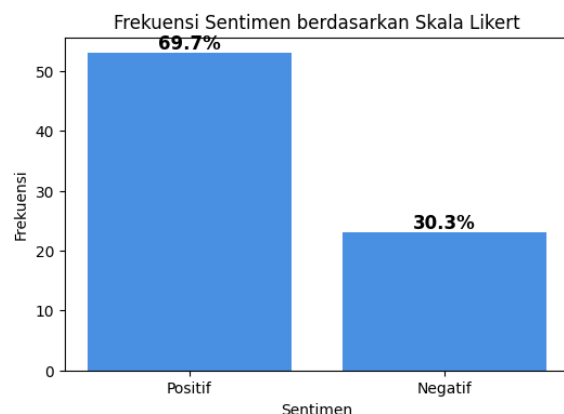
Merupakan proses menghilangkan kata sambung yang terdapat pada dataset. Merupakan proses penyaringan kata pada dataset yang tidak ada hubungan nya dalam sentimen analisis. Berikut merupakan contoh hasil dari *stopwords*.

Tabel 3. Stopwords

Data Mentah	Data Setelah Stopwords
'lebih', 'positif', 'menurutku', 'karena', 'bisa', 'mengefisiensi', 'waktu', 'pengerjaan', 'tapi', 'ya', 'bisa', 'jadi', 'negatif', 'karena', 'terlalu', 'bergantung'.	'lebih', 'positif', 'turut', 'karena', 'efisiensi', 'waktu', 'kerja', 'tapi', 'bisa', 'negatif', 'karena', 'gantung'.

4.7. Labelling

Setelah dilakukan pengambilan proses pelabelan sentimen ini dilakukan menggunakan dua pendekatan, yaitu menggunakan kualitatif melalui 20 jenis pertanyaan skala likert untuk mengukur tingkat kepuasan responden terhadap penggunaan AI bagi pendidikan. Dengan rentang nilai skala 1 hingga 5, nilai 1-2 mengindikasikan ketidaksetujuan atau cenderung tidak terbantu dengan adanya AI dan condong ke negatif. Namun jika rentang 3 maka bernilai netral. Sedangkan, jika rentang nilai 4-5 menunjukkan kepuasan maksimal dengan adanya AI yang condong kategori positif. Dapat dilihat dari Gambar 6.



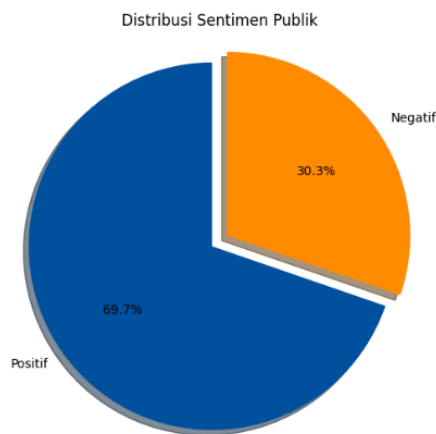
Gambar 6. Distribusi klasifikasi kategori sentimen

Berdasarkan hasil dari pelabelan, diperoleh 69,7% sentimen positif dan sentimen negatif sebanyak 30,3%. hal ini menunjukkan mayoritas responden memiliki persepsi yang cenderung positif.

Selain itu, terdapat satu jenis pertanyaan observasi yang digunakan untuk memberi wadah bagi responden untuk memberikan opini mereka lebih mendalam terhadap pernggunaan AI bagi pendidikan dimasa depan, serta berbagai kekhawatiran terhadap etika-etika penggunaan AI agar tidak melanggar ketentuan akademik. Beberapa responden mengungkapkan pendapat positif mereka terhadap AI yang membantu meningkatkan pemahaman materi lebih mendalam, brainstorming kalimat-kalimat baru, serta penelitian menjadi lebih cepat. Di sisi lain, beberapa responden menyatakan kekhawatiran

mengenai penggunaan AI dibidang pendidikan dapat menyebabkan ketergantungan mahasiswa dengan mengandalkan AI sehingga dapat mengurangi kemampuan berpikir kritis dan tingginya tingkat plagiat.

Dalam analisis sentimen ini, proses pelabelan data dilakukan dengan menggunakan python dengan Pustaka *TextBlob*. Hasil evaluasi awal menunjukkan hasil dari sentimen mahasiswa terhadap penggunaan AI terhadap kreativitas dan efektivitas dalam menyelesaikan tugas akhir lebih condong ke kategori positif dibandingkan kategori sentimen negatif. Dapat dilihat data pada grafik dibawah ini.



Gambar 7. Visualisasi sentimen

Dari grafik tersebut, menunjukkan bahwa sentimen positif paling dominan, sebanyak 69,7% ulasan, dan sentimen negatif sebanyak 30,3%. Namun memiliki nilai accuracy yang cukup kecil sehingga menjadi pertimbangan untuk menggunakan model Decision tree ini, diperlukan perbaikan dengan menambahkan metode SMOTE untuk mengatasi data yang tidak seimbang. Teknik ini membantu menyeimbangkan kelas yang tidak seimbang dalam dataset, selain itu, dilakukan pelabelan ulang dengan menambahkan definisi daftar kata-kata positif dan negative, yang bertujuan untuk mengklasifikasikan teks secara lebih akurat.

4.8. Pembobotan Kata *TF-IDF*

Setelah dilakukan proses *preprocessing*, Langkah selanjutnya adalah teks akan dilakukan pemberian bobot kata menggunakan *TF-IDF* untuk menghitung kata-kata yang memiliki bobot tinggi, serta mengubah teks menjadi data numerik agar lebih sederhana dan mudah diolah.

DataFrame TF-IDF:									
	ai	atau	bagi	banyak	batas	bisa	contoh	\	
0	0.000000	0.210225	0.000000	0.210225	0.000000	0.159082	0.210225		
1	0.000000	0.000000	0.000000	0.000000	0.000000	0.403566	0.000000		
2	0.271379	0.000000	0.271379	0.000000	0.271379	0.000000	0.000000		
	dampak	depan	efisiensi	...	sangat	saya	sekitar	tambah	
0	0.000000	0.000000	0.000000	...	0.000000	0.42045	0.210225	0.210225	
1	0.000000	0.000000	0.26532	...	0.000000	0.000000	0.000000	0.000000	
2	0.271379	0.271379	0.000000	...	0.271379	0.000000	0.000000	0.000000	

Gambar 8. Pembobotan tf-idf

Dengan transformasi ini, *TF-IDF* membantu model memprioritaskan kata-kata yang lebih relevan, sehingga mendukung proses klasifikasi sentimen dengan hasil yang lebih akurat. Proses ini penting dalam menganalisis data teks, karena mengubah teks mentah menjadi format yang dapat dipahami dan diolah oleh algoritma machine learning.

4.9. Klasifikasi Decision Tree

Setelah semua kata diberikan bobot nilai, tahap selanjutnya yaitu melakukan pelatihan menggunakan model *decision tree*. Dataset dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian akurasi model. Berguna untuk memastikan model memiliki cukup data untuk belajar serta diuji dengan data yang belum diketahui sebelumnya. Selama tahap pelatihan, model akan menghitung probabilitas kata terbesar untuk setiap sentimen. Kemudian digunakan untuk memprediksi kategori sentimen dari data yang diuji.

```
from sklearn.model_selection import train_test_split
# Membagi data menjadi data latih dan data uji (61 data latih, 15 data uji)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=15/76, random_state=42)

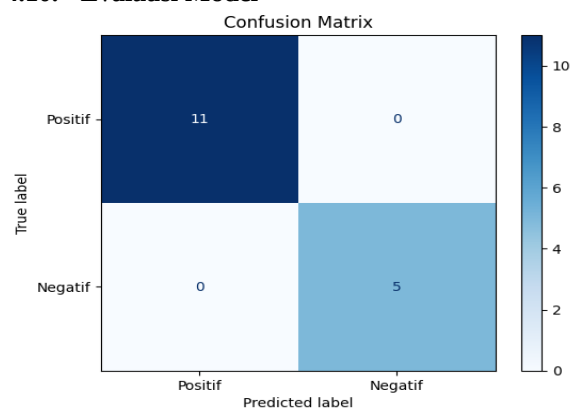
vectorizer = TfidfVectorizer(stop_words=stopwords.words('indonesian'))
X_train_tfidf = vectorizer.fit_transform(X_train)
X_test_tfidf = vectorizer.transform(X_test)

print("X Train:", X_train)
print("Y Train:", y_train)
print("X Test:", X_test)
print("Y Test:", y_test)
```

Gambar 9. Split data

Setelah membagi data dan menerapkan TF-IDF, langkah berikutnya adalah melatih model menggunakan data latih (X_{train_tfidf} , y_{train}) dan mengevaluasi performa model pada data uji (X_{test_tfidf} , y_{test}). Proses ini bertujuan mengukur akurasi dan kemampuan klasifikasi sentimen.

4.10. Evaluasi Model



Gambar 10. Confusion matrix

Accuracy: 0.44				
Classification Report:				
	precision	recall	f1-score	support
0	0.60	0.55	0.57	11
1	0.17	0.20	0.18	5
accuracy			0.44	16
macro avg	0.38	0.37	0.38	16
weighted avg	0.46	0.44	0.45	16

Gambar 11. Akurasi sebelum proses SMOTE

Pada tahap awal, pelatihan model *decision tree* tanpa dilakukan penyeimbangan antar kelas, menghasilkan akurasi sebesar 44%. Rendahnya akurasi ini disebabkan oleh ketidakseimbangan jumlah data pada masing-masing kelas sentimen, sehingga terjadinya bias dalam dataset. Untuk mengatasi hal ini diperlukan *SMOTE* algoritma tambahan. Selain itu, diperlukan untuk membuat definisi kata-kata positif dan negatif yang berfungsi sebagai acuan dalam proses pelabelan sentimen. Dapat dilihat dari gambar dibawah.

```
# Definisi kata-kata positif dan negatif
positive_words = ['positif', 'membantu', 'meningkatkan', 'efektif', 'kreatif', 'baik']
negative_words = ['negatif', 'bermasalah', 'kekhawatiran', 'terlalu bergantung', 'pe']

# Fungsi untuk mengklasifikasi berdasarkan jumlah kata positif dan negatif
def classify_sentiment_enhanced(text):
    text = text.lower()
    positive_count = sum(1 for word in positive_words if word in text)
    negative_count = sum(1 for word in negative_words if word in text)
```

Gambar 12. Tahap pendefinisian daftar kata

Setelah proses penyeimbangan data dan penerapan *SMOTE*, akurasi model meningkat signifikan menjadi 88%, di mana nilai precision, recall, dan F1-score untuk kelas positif mencapai 91%, sedangkan untuk kelas negatif sedikit lebih rendah dengan nilai 80%. nilai rata-rata macro dan weighted masing-masing tercatat sebesar 85 % dan 88%, yang menunjukkan keseimbangan yang baik antara kedua kelas, seperti yang dapat dilihat pada Gambar 13.

Accuracy: 0.88				
Classification Report:				
	precision	recall	f1-score	support
0	0.91	0.91	0.91	11
1	0.80	0.80	0.80	5
accuracy			0.88	16
macro avg	0.85	0.85	0.85	16
weighted avg	0.88	0.88	0.88	16

Gambar 13. Akurasi setelah proses SMOTE

Kemudian data tersebut akan ditampilkan menggunakan visualisasi word cloud memperlihatkan kata dengan frekuensi terbanyak atau kata yang sering muncul dalam dataset opini. Semakin banyak kata tersebut muncul maka semakin menonjol dibanding kata yang lain. Dapat dilihat kata positif, ai, membantu sebagai kata yang paling dominan serta menunjukkan bahwa mayoritas kalimat mengarah kategori sentimen positif. Peningkatan pada akurasi ini mengindikasikan bahwa perpaduan antara penyeimbangan data dan pembobotan kata berkontribusi secara signifikan terhadap peningkatan performa model dalam mengklasifikasikan sentimen analisis.



Gambar 14. Wordcloud

5. KESIMPULAN DAN SARAN

Penelitian ini menggunakan algoritma Decision Tree untuk melakukan Analisis Sentimen Peran *Artificial Intelligence* (AI) Terhadap Kreativitas Dan Efektivitas Mahasiswa Dalam Menyelesaikan Tugas Akhir Menggunakan *Decision Tree* Berbasis *SMOTE*. Dikarenakan algoritma ini mudah dimengerti dalam proses klasifikasi. Setelah dilakukan pengumpulan data dari penyebaran kuesioner diperoleh 150 data responden. Setelah melakukan *pre-processing* data, didapatkan 76 data yang siap untuk dianalisis sesuai dengan perhitungan rumus slovin untuk tahap pengambilan sampel populasi. Data kemudian akan dibagi dengan proposi 80% *data training* dan 20% *data testing*. Setelah dilakukan analisis menggunakan algoritma *decision tree* memperoleh akurasi hanya 44%, dapat dikatakan sangat rendah sehingga model tersebut belum efektif untuk digunakan pada penelitian ini. Untuk meningkatkan kinerjanya, diterapkan metode *Synthetic Minority Over-Sampling Technique* (*SMOTE*), sehingga menghasilkan akurasi 88%, nilai precision, recall, dan F1-score untuk kelas positif mencapai 91%, sedangkan untuk kelas negatif sedikit lebih rendah dengan nilai 80%. nilai rata-rata macro dan weighted masing-masing tercatat sebesar 85% dan 88%, yang menandakan keakuratan dalam klasifikasi sentimen.

Hasil klasifikasi sentimen menunjukkan bahwa mayoritas sentimen yang terkandung dalam opini mahasiswa mengenai penggunaan AI untuk menyelesaikan tugas akhir cenderung lebih banyak positif, yaitu sebanyak 69,7% sementara sisanya, yaitu 30,3% tergolong dalam kategori negatif. Berdasarkan hasil penelitian ini, rekomendasi untuk penelitian selanjutnya adalah mengumpulkan data yang lebih banyak dengan melibatkan mahasiswa dari berbagai jurusan, bukan hanya dari jurusan Teknik di Universitas Pelita Bangsa. Dengan jumlah data yang lebih banyak, proses pembagian data dapat memberikan hasil yang lebih optimal dan meningkatkan kinerja model algoritma.

DAFTAR PUSTAKA

- [1] A. Kaharudin, A. A. Supriyadi, Muhlis, H. Baitika, and M. Derryanur, "Analisis Sentimen pada Media Sosial dengan Teknik Kecerdasan Buatan Naïve Bayes: Kajian Literatur Review," *OKTAL J. Ilmu Komput. Dan Sains*, vol. 2, no. 06, Art. no. 06, Jun. 2023.
- [2] Z. Wang, Q. Xie, Y. Feng, Z. Ding, Z. Yang, and R. Xia, "Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study," Feb. 17, 2024, arXiv: arXiv:2304.04339. doi: 10.48550/arXiv.2304.04339.
- [3] A. Fatkhudin, F. A. Artanto, N. A. Safli, and D. Wibowo, "Decision Tree Berbasis SMOTE Dalam Analisis Sentimen Penggunaan Artificial Intelligence Untuk Skripsi," *REMIK Ris. Dan E-J. Manaj. Inform. Komput.*, vol. 8, no. 2, Art. no. 2, Apr. 2024, doi: 10.33395/remik.v8i2.13531.

- [4] C. Cahyaningtyas, Y. Nataliani, and I. R. Widiyari, "Analisis Sentimen Pada Rating Aplikasi Shopee Menggunakan Metode Decision Tree Berbasis SMOTE," *AITI*, vol. 18, no. 2, Art. no. 2, Nov. 2021, doi: 10.24246/aiti.v18i2.173-184.
- [5] D. Dablain, B. Krawczyk, and N. V. Chawla, "DeepSMOTE: Fusing Deep Learning and SMOTE for Imbalanced Data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 9, pp. 6390–6404, Sep. 2023, doi: 10.1109/TNNLS.2021.3136503.
- [6] K. Luthfiyyah, L. Zhafira, S. Nurani, and S. F. Giwangsa, "Analisis Peran Artificial Intelligence (AI): ChatGPT dalam Perkuliahan di Kalangan Mahasiswa PGSD Universitas Pendidikan Indonesia," vol. 8, 2024.
- [7] R. Vincent, I. Maulana, and O. Komarudin, "PERBANDINGAN KLASIFIKASI NAIVE BAYES DAN SUPPORT VECTOR MACHINE DALAM ANALISIS SENTIMEN DENGAN MULTICLASS DI TWITTER," *JATI J. Mhs. Tek. Inform.*, vol. 7, no. 4, Art. no. 4, 2023, doi: 10.36040/jati.v7i4.7152.
- [8] A. Azrul, A. I. Purnamasari, and I. Ali, "ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP PERKEMBANGAN ARTIFICIAL INTELLIGENCE DENGAN PENERAPAN ALGORITMA LONG SHORT-TERM MEMORY (LSTM)," *JATI J. Mhs. Tek. Inform.*, vol. 8, no. 1, Art. no. 1, Feb. 2024, doi: 10.36040/jati.v8i1.8416.
- [9] K. K. Saфра and E. Zuliarso, "ANALISIS SENTIMEN TERHADAP PELAKSANAAN PILKADA 2024 PADA MEDIA SOSIAL YOUTUBE MENGGUNAKAN METODE DECISION TREE," *J. Inform. Teknol. Dan Sains Jinteks*, vol. 7, no. 1, Art. no. 1, Feb. 2025, doi: 10.51401/jinteks.v7i1.5295.
- [10] B. Wicaksono and N. Cahyono, "ANALISIS SENTIMEN KOMENTAR INSTAGRAM PADA PROGRAM KAMPUS MERDEKA DENGAN ALGORITMA NAIVE BAYES DAN DECISION TREE," *JATI J. Mhs. Tek. Inform.*, vol. 8, no. 2, Art. no. 2, Apr. 2024, doi: 10.36040/jati.v8i2.9473.
- [11] U. D. Akor, M. R. F. Payu, and L. O. Nashar, "Penerapan Metode I-CHAID Menggunakan SMOTE pada Data Tidak Seimbang untuk Klasifikasi Durasi Studi Mahasiswa," *Jambura J. Math.*, vol. 7, no. 1, Art. no. 1, Feb. 2025, doi: 10.37905/jjom.v7i1.27978.
- [12] U. Lutfiani, "EFFECT OF CREATIVITY ON STUDENT LEARNING ACHIEVEMENT," *Psikoeduko J. Psikol. Edukasi Dan Konseling*, vol. 1, no. 1, Art. no. 1, Mar. 2021, doi: 10.17509/psikoeduko.v1i1.32827.
- [13] S. Suyuti, P. M. E. Wahyuningrum, M. A. Jamil, M. L. Nawawi, D. Aditia, and N. G. A. L. Rusmayani, "Analisis Efektivitas Penggunaan Teknologi dalam Pendidikan Terhadap Peningkatan Hasil Belajar," *J. Educ.*, vol. 6, no. 1, Art. no. 1, May 2023, doi: 10.31004/joe.v6i1.2908.
- [14] R. G. Guntara, "Pemanfaatan Google Colab Untuk Aplikasi Pendeteksian Masker Wajah Menggunakan Algoritma Deep Learning YOLOv7," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 5, no. 1, Art. no. 1, Feb. 2023, doi: 10.47233/jteksis.v5i1.750.
- [15] K. C. Astuti, A. Firmansyah, and A. Riyadi, "Implementasi Text Mining Untuk Analisis Sentimen Masyarakat Terhadap Ulasan Aplikasi Digital Korlantas Polri pada Google Play Store," *REMIK Ris. Dan E-J. Manaj. Inform. Komput.*, vol. 8, no. 1, Art. no. 1, Jan. 2024, doi: 10.33395/remik.v8i1.13421.
- [16] M. H. Mahendra, D. T. Mardiansyah, and K. M. Lhaksana, "Analisis Sentimen Tweet COVID-19 Menggunakan Metode K-Nearest Neighbors dengan Ekstraksi Fitur TF-IDF dan CountVectorizer," vol. 1, 2023.
- [17] E. Aditya, I. G. K. Astawa, K. G. Limbong, G. Indrawan, and M. A. O. Gunawan, "Analisis Sentimen Pengguna Sistem E-Kinerja Desa Kabupaten Jembrana Menggunakan Metode Naive Bayes," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 7, no. 1, Art. no. 1, Jan. 2025, doi: 10.47233/jteksis.v7i1.1693.
- [18] M. Danny, A. Muhidin, and A. Jamal, "Application of the K-Nearest Neighbor Machine Learning Algorithm to Product Sales of Best-Selling Products," *Brill. Res. Artif. Intell.*, vol. 4, no. 1, pp. 255–264, Jun. 2024, doi: 10.47709/brilliance.v4i1.4063.
- [19] A. Azis, A. T. Zy, and A. S. Sunge, "Prediksi Penjualan Obat Dan Alat Kesehatan Terlaris Menggunakan Algoritma K-Nearest Neighbor," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 1, pp. 117–124, Jan. 2024, doi: 10.47233/jteksis.v6i1.1078.
- [20] R. A. Firsttama, A. A. Arifiyanti, and D. S. Y. Kartika, "Analisis Sentimen Komentar Youtube Konferensi Tingkat Tinggi G20 Menggunakan Metode Naive Bayes," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 2, Art. no. 2, Apr. 2024, doi: 10.47233/jteksis.v6i2.1263.