ANALYSIS OF DIFFERENCES BETWEEN AI AND HUMAN TEXTS USING THE NATURAL LANGUAGE PROCESSING METHOD

ANALISIS PERBEDAAN TEKS BUATAN ARTIFICIAL INTELLIGENCE DAN MANUSIA MENGGUNAKAN METODE NATURAL LANGUAGE PROCESSING

Dinda Cahyana[1], VitoReyLukito Sijabat[2], Mohammad Irfan Fahmi[3]
[1,2,3]Universitas Prima Indonesia, Jl. Sampul No.3, Kota Medan, Sumatera Utara, Indonesia
email: dndachyna@icloud.com[1], vitoreylukitos@gmail.com[2], mohammadirfanfahmi@unprimdn.ac.id[3]

Abstract *- Artificial Intelligence has become increasingly proficient in generating text that mimics human writing, yet existing detection tools remain limited in accuracy and adaptability. Previous studies indicate that systems like Turnitin and GPTZero often perform below 80% accuracy and struggle with paraphrased or advanced AI-generated content. This study addresses that gap by analyzing linguistic differences between AI-generated and human-written texts using Natural Language Processing. A dataset of 487,235 texts (305,797 human-written and 181,438 AI-generated) was processed using TF-IDF vectorization and classified with the Multinomial Naive Bayes algorithm. The model achieved 99.35% accuracy and an F1-score of 0.9948, with balanced performance in detecting both text types. Results show that while AI-generated texts are structurally consistent, they often lack the emotional depth and cultural nuance found in human writing. These findings suggest NLP methods are highly effective in distinguishing between the two, and have practical implications for developing more reliable detection systems to ensure textual authenticity in education, journalism, and digital media monitoring.*

Keywords *- generative text, artificial intelligence, human writing, NLP, linguistic study.*

Abstrak Kecerdasan buatan kini semakin mampu menghasilkan teks yang menyerupai tulisan manusia, namun alat deteksi yang ada masih terbatas dalam hal akurasi dan fleksibilitas. Studi sebelumnya menunjukkan bahwa alat seperti Turnitin dan GPTZero sering memiliki akurasi di bawah 80% dan kesulitan mengenali teks hasil parafrase atau buatan model AI canggih. Penelitian ini mengisi celah tersebut dengan menganalisis perbedaan linguistik antara teks AI dan teks manusia menggunakan metode Natural Language Processing. Sebanyak 487.235 teks (305.797 teks manusia dan 181.438 teks AI) diproses menggunakan TF-IDF dan diklasifikasikan dengan algoritma Multinomial Naive Bayes. Model berhasil mencapai akurasi 99,35% dan F1-score 0,9948, dengan kinerja seimbang dalam mengenali kedua jenis teks. Hasil menunjukkan bahwa meskipun teks AI terstruktur rapi, teks tersebut cenderung kurang ekspresif dan tidak menangkap konteks budaya seperti tulisan manusia. Temuan ini menunjukkan bahwa metode NLP sangat efektif untuk membedakan kedua jenis teks, dan memiliki implikasi penting dalam pengembangan sistem deteksi otomatis untuk menjamin keaslian teks dalam pendidikan, jurnalistik, dan pengawasan konten digital.

Kata Kunci - teks generatif, kecerdasan buatan, tulisan manusia, NLP, studi linguistic.

I. PENDAHULUAN

Di zaman sekarang yang serba digital, kecerdasan buatan (AI) sudah banyak dipakai di berbagai bidang, termasuk untuk membuat teks atau tulisan[1]. Teknologi seperti GPT, BERT, dan Transformer—yang merupakan bagian dari pemrosesan bahasa alami (NLP) telah terbukti sangat bagus dalam membuat teks yang mirip seperti tulisan manusia Dengan kemampuan ini, AI telah digunakan untuk menghasilkan berbagai jenis tulisan, mulai dari artikel berita, konten pemasaran, hingga karya ilmiah. Namun, meskipun AI bisa meniru gaya dan struktur tulisan manusia, sebenarnya masih ada perbedaan yang bisa ditemukan, seperti dalam susunan kalimat, makna, dan pola Bahasa. Seiring meningkatnya kecanggihan teknologi NLP, perbedaan antara teks buatan manusia dan AI menjadi semakin sulit dikenali. Salah satu tantangan utama dalam penelitian ini adalah merancang metode yang lebih akurat untuk mengidentifikasi teks yang dihasilkan oleh AI [3]. Walaupun sejumlah alat pendeteksi telah tersedia, sebagian besar masih memiliki keterbatasan akurasi dan rentan terhadap manipulasi oleh model AI yang semakin canggih.

Penelitian sebelumnya menunjukkan bahwa teks yang dihasilkan AI cenderung bersifat kaku dan kurang mencerminkan dinamika linguistik alami seperti yang biasa ditemukan dalam tulisan manusia, meskipun secara sekilas sering kali sulit dibedakan. Hal ini terjadi karena model AI masih memiliki keterbatasan dalam menangkap konteks yang kompleks serta ekspresi bahasa yang beragam, yang menjadi ciri khas tulisan manusia[4]. Temuan dalam penelitian ini menunjukkan adanya peluang untuk mengembangkan pendekatan yang lebih presisi dan komprehensif dalam membedakan antara teks buatan AI dan teks manusia, melalui pemanfaatan teknik NLP yang lebih canggih dan mutakhir.

Berdasarkan landasan teori yang telah dijelaskan, penelitian ini bertujuan untuk mengembangkan model deteksi teks buatan AI yang lebih akurat dan berimbang, serta memperdalam pemahaman terhadap perbedaan linguistik antara teks yang dihasilkan oleh manusia dan oleh AI[5]. Sebagai upaya untuk permasalahan ini, penelitian ini akan menggunakan berbagai metode NLP guna menganalisis karakteristik linguistik dari kedua jenis teks. Pendekatan yang digunakan mencakup teknik ekstraksi dan analisis statistik, yang bertujuan untuk mengevaluasi performa model melalui metrik akurasi, precision, recall, dan F1-score.

Penelitian ini bertujuan untuk mengidentifikasi dan menganalisis perbedaan linguistik linguistik antara teks yang ditulis oleh manusia dan yang dihasilkan oleh model AI, khususnya GPT dan BERT, dengan memanfaatkan teknik NLP terkini. Fokus utama penelitian ini adalah untuk mengeksplorasi karakteristik linguistik yang membedakan kedua jenis teks tersebut, termasuk CountVectorizer, TfidTransformer, dan MultinomialNB [6]. Oleh karena itu, diperlukan pengembangan pendekatan yang lebih tepat dan menyeluruh guna membedakan secara efektif antara teks buatan manusia dan teks buatan AI[7].

II. SIGNIFIKASI STUDI

A. Studi Literatur

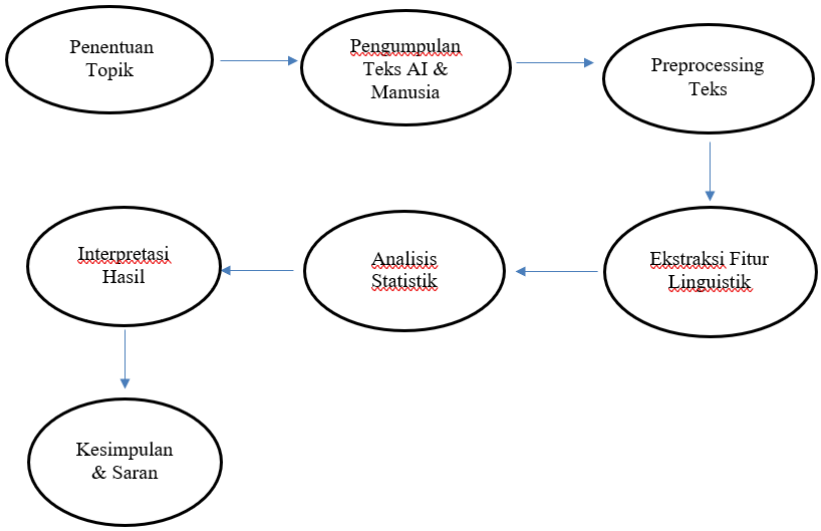
Seiring dengan kemajuan teknologi NLP, banyak penelitian telah dilakukan untuk membedakan teks yang dihasilkan oleh manusia dan oleh AI, khususnya model bahasa besar seperti GPT dan BERT[8]. Berbagai pendekatan telah digunakan, mulai dari analisis linguistik tradisional hingga penerapan model pembelajaran mesin. Studi-studi terdahulu memberikan landasan penting bagi penelitian ini, namun masih menyisakan celah yang dapat dijelajahi lebih lanjut[9]. Salah satu studi yang menonjol adalah penelitian oleh Ippolito et al. (2020) yang menguji kemampuan manusia dalam mengidentifikasi teks buatan GPT-2. Hasil penelitian mereka menunjukkan bahwa manusia hanya sedikit lebih baik dari peluang acak dalam mengenali teks buatan AI, menandakan bahwa kualitas teks AI semakin menyerupai teks manusia. Penelitian ini berfokus pada evaluasi persepsi manusia, tetapi tidak secara mendalam membedah ciri-ciri linguistik yang membedakan teks tersebut[10].

Penelitian lain oleh Bakhtin et al. (2019) menggunakan pendekatan berbasis klasifikasi machine learning untuk membedakan teks manusia dan AI, dengan memanfaatkan fitur seperti panjang kalimat, frekuensi kata, dan skor perplexity. Meskipun pendekatan ini efektif dalam kondisi tertentu, model deteksi yang dihasilkan cenderung sensitif terhadap perubahan model AI dan mudah dikalahkan oleh model AI generasi baru[11]. Sementara itu, Gehrmann et al. (2019) mengembangkan alat deteksi yang disebut GLTR (Giant Language model Test Room), yang memanfaatkan probabilitas prediksi kata dari model bahasa untuk mendeteksi teks AI. GLTR mengandalkan hipotesis bahwa teks buatan AI cenderung menggunakan kata-kata dengan probabilitas tinggi secara konsisten, sedangkan tulisan manusia lebih bervariasi. Meskipun pendekatan ini memberikan visualisasi yang bermanfaat, metode ini masih kurang akurat ketika digunakan terhadap model AI yang lebih canggih seperti GPT-3 atau GPT-4[12].

Di sisi lain, penelitian oleh Zellers et al. (2019) dalam proyek Grover menunjukkan bahwa model detektor yang dilatih menggunakan teks buatan oleh model tertentu akan cenderung bekerja lebih baik pada model yang sama, tetapi gagal mendeteksi teks dari model lain. Hal ini menunjukkan keterbatasan generalisasi metode yang berbasis model tertentu[13]. Penelitian ini menggunakan Natural Language Processing untuk menganalisis teks buatan AI dan teks manusia. Data yang digunakan dalam penelitian ini berasal dari dataset gabungan yang terdiri atas teks buatan manusia dan teks yang dihasilkan oleh AI. Jumlah total data adalah 487.235 teks, dengan rincian 305.797 teks ditulis oleh manusia dan 181.438 teks dihasilkan oleh AI. Data disimpan dalam format CSV dan dimuat menggunakan library *pandas* di Python. Pada tahap ini, jumlah baris dan kolom dihitung, dan kolom yang kosong dihapus untuk memastikan kualitas data. Langkah selanjutnya adalah proses pembersihan dan normalisasi data teks (cleaning & preprocessing). Ini meliputi pengubahan seluruh huruf menjadi huruf kecil, penghapusan tautan (link), tanda baca, angka, dan spasi berlebih. Teks yang telah dibersihkan kemudian diubah menjadi representasi numerik menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Proses ini dilakukan untuk mengidentifikasi kata-kata yang penting dan sering muncul dalam dokumen, serta membentuk vektor fitur sebagai input untuk algoritma klasifikasi. Hasil dari TF-IDF digunakan sebagai dasar pelatihan model klasifikasi teks. Hasil evaluasi menunjukkan bahwa model mampu membedakan antara teks AI dan teks manusia dengan akurasi sebesar 99,35% dan rata-rata F1-score sebesar 0,99. Ini menandakan bahwa model sangat efektif dalam mendeteksi dan mengklasifikasikan kedua jenis teks tersebut. Hal ini juga menjadikan NLP sebagai metode yang sangat direkomendasikan untuk menguji perbedaan teks AI dan teks manusia.

B. Metode Penelitian

Untuk memastikan penelitian ini dapat berjalan secara sistematis dan selesai tepat waktu, peneliti menyusun sebuah diagram alur penelitian. Diagram tersebut dapat dilihat pada gambar 1.



Gambar 1. Alur Penelian.

1. Penentuan Topik

Tahap awal adalah menentukan topik penelitian, yaitu membandingkan teks buatan AI dengan teks buatan manusia menggunakan pendekatan NLP. Peneliti juga menetapkan tujuan, rumusan masalah, dan ruang lingkup penelitian.

2. Pengumpulan Data

Dalam tahap awal pemrosesan data, peneliti menggunakan Google Colab untuk menjalankan skrip Python secara daring. Library Python pantas digunakan untuk membaca dan mengelola dataset yang telah disimpan dalam format CSV[14]. Pada tahap awal, peneliti menentukan path file CSV yang berisi data teks yang akan dianalisis, kemudian memuat data tersebut ke dalam sebuah DataFrame menggunakan fungsi pd.read_csv(). Setelah proses pemuatan data, peneliti menampilkan lima entri pertama dari dataset dengan menggunakan fungsi df.head() serta menampilkan daftar nama kolom yang terdapat dalam dataset. Dataset yang digunakan terdiri dari dua kolom utama, kolom "text", yang memuat isi ulasan, dan kolom "generated", yang menunjukkan label apakah teks tersebut ditulis oleh manusia atau dihasilkan oleh model AI.

```
First 5 records:
                                                text  generated
0  Cars. Cars have been around since they became ...        0.0
1  Transportation is a large necessity in most co...        0.0
2  "America's love affair with it's vehicles seem...        0.0
3  How often do you ride in a car? Do you drive a...        0.0
4  Cars are a wonderful thing. They are perhaps o...        0.0

Kolom-kolom: Index(['text', 'generated'], dtype='object')
```

Gambar 2. Dataset yang ingin diolah

Untuk mengetahui proporsi data dalam dataset, peneliti melakukan perhitungan terhadap total jumlah teks serta membedakan antara teks yang ditulis oleh manusia dan yang dihasilkan oleh AI:

Tabel 1. Proporsi Data

Total teks	487.235
Teks buatan manusia	305.797
Teks hasil AI	181.438

Langkah ini menjadi dasar untuk analisis lanjutan, seperti pelabelan, pemisahan data, atau pelatihan model klasifikasi teks.

3. *Pra-Pemrosesan Data*

Data teks yang telah dikumpulkan dilakukan proses melalui tahap preprocessing atau pra-pemrosesan yaitu penghapusan tanda baca, stopword removal, stemming/lemmatisasi, dan normalisasi teks untuk memastikan data siap dianalisis secara linguistik.

- Data Cleaning, dilakukan dengan membersihkan kolom teks menggunakan fungsi clean_text. Proses pembersihan ini mencakup konversi seluruh teks menjadi huruf kecil, menghapus tautan, tanda baca, angka, serta spasi yang berlebih. Tahapan ini dilakukan guna memastikan bahwa bahwa data yang dimasukkan ke dalam model *machine learning* dalam kondisi yang bersih dan siap untuk dianalisis secara efisien.

```
[ ]  import re
     import string

     def clean_text(text):
         text = text.lower()
         text = re.sub(r"http\S+", "", text)  # Hapus link
         text = re.sub(rf"[{string.punctuation}]", "", text)  # Hapus tanda baca
         text = re.sub(r"\d+", "", text)  # Hapus angka
         text = re.sub(r"\s+", " ", text).strip()  # Normalisasi spasi
         return text

     # Terapkan ke kolom teks
     df['clean_text'] = df['text'].astype(str).apply(clean_text)
```

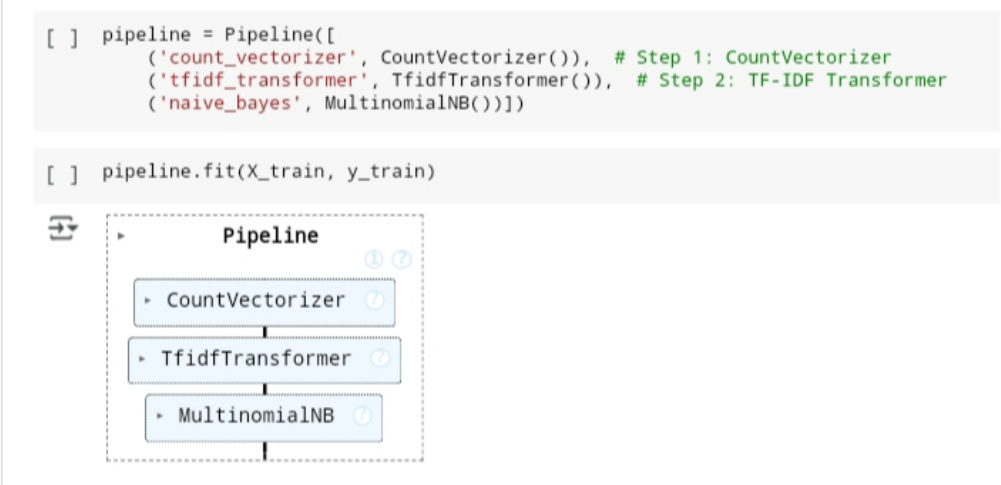
Gambar 3. Penghapusan link, tanda baca, angka, dan spasi

- Stop Word Removal, Pada tahap ini, dilakukan penghapusan *stopwords*, yaitu kata-kata yang sering muncul namun memiliki nilai informasi yang rendah dalam analisis, seperti "yang", "di", "itu", dan kata sejenis lainnya. Langkah ini bertujuan untuk meningkatkan kualitas analisis dengan mengurangi elemen linguistik yang bersifat umum dan kurang relevan
- Stemming, Proses mengubah kata-kata ke bentuk kata dasarnya dengan memotong imbuhan. Sebagai contoh, kata "pembelajaran" diubah menjadi bentuk dasarnya, yaitu "belajar". Tujuan dari proses ini adalah untuk menyamakan kata-kata yang memiliki makna serupa, sehingga dapat memperkuat konsistensi data. Dengan demikian, teks yang digunakan dalam analisis akan lebih bersih dan bebas dari gangguan elemen non-linguistik yang dapat memengaruhi hasil.

4. *Ekstraksi Fitur Linguistik*

Setelah proses pembersihan data selesai dilakukan, langkah berikutnya adalah melakukan ekstraksi fitur linguistik dari kedua kategori teks. Fitur-fitur ini mencerminkan karakteristik kebahasaan yang khas dan digunakan untuk menganalisis perbedaan antara teks yang dihasilkan oleh AI dan teks yang ditulis oleh manusia dalam konteks penerapan metode Natural Language Processing (NLP [15]. Proses ekstraksi fitur linguistik dimulai dengan membentuk kosakata dan menghitung bobot setiap kata. Tahap awal menggunakan

CountVectorizer untuk mengonversi teks menjadi matriks frekuensi kata, lalu dilanjutkan dengan TfidfTransformer guna menghitung skor TF-IDF yang menunjukkan tingkat kepentingan suatu kata dalam keseluruhan dokumen. [16]. Proses ini membentuk dasar dari pipeline pemrosesan teks yang digunakan. Berikut adalah pipeline yang digunakan:



Gambar 4. Ekstraksi Fitur Linguistik

Dengan pendekatan ini mengubah kata-kata dalam teks menjadi data numerik yang mencerminkan makna dan pola distribusinya, yang kemudian digunakan sebagai input untuk melatih model klasifikasi teks seperti Multinomial Naive Bayes.

5. *Analisis Statistik*
Fitur-fitur yang diestrak dianalisis secara statistik (precision, recall, F1-score, support) mengidentifikasi adanya perbedaan yang signifikan antara kedua jenis teks.

Rumus F1-score:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} \times \text{Recall}} \quad (1)$$

Precision:

$$\frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Positive\ (FP)} \quad (2)$$

Recall:

$$\frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Negative\ (NP)} \quad (3)$$

Dimana:

Precision = TP / (TP + FP)

(Berapa banyak dari prediksi positif yang benar)

Recall = TP / (TP + FN)

(Berapa banyak dari prediksi positif yang berhasil di temukan)

TP = True Positive

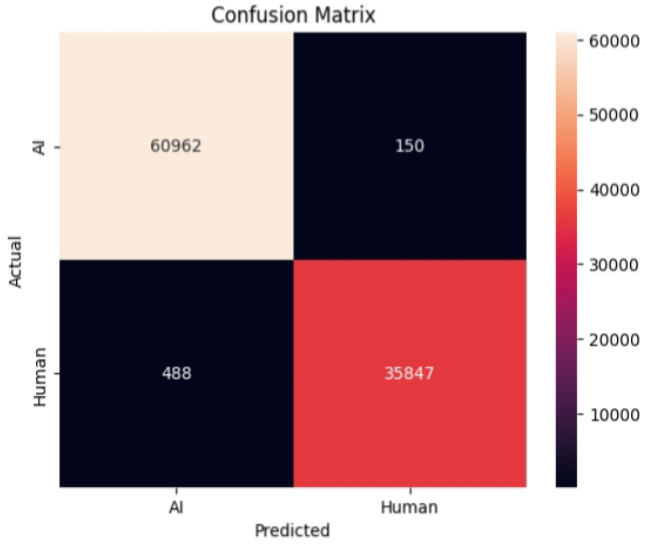
FP = False Positive

FN = False Negative

Berdasarkan hasil analisis statistik yang diperoleh, dapat disimpulkan bahwa model yang digunakan memiliki tingkat akurasi yang tinggi dalam membedakan antara teks buatan AI dan teks yang ditulis oleh manusia. Hal ini menunjukkan bahwa, meskipun model generatif seperti AI semakin canggih, tetap terdapat ciri linguistik khas yang membedakan keduanya. Dalam konteks Natural Language Processing (NLP), hasil ini memberikan kontribusi signifikan, khususnya dalam verifikasi keaslian teks, deteksi otomatis terhadap konten yang dihasilkan oleh AI, serta pengembangan sistem yang mampu beradaptasi dengan beragam gaya bahasa alami manusia.

III. HASIL DAN PEMBAHASAN

Hasil evaluasi dari model klasifikasi disajikan dalam bentuk *confusion matrix*, yang mencakup sejumlah metrik evaluasi penting, seperti akurasi, presisi, *recall*, dan *F1-score*. Matriks ini memberikan gambaran yang lebih mendalam khususnya dalam mengidentifikasi dan membedakan antara teks yang ditulis oleh manusia dan teks yang dihasilkan oleh sistem generatif berbasis AI [17] Confusion matrix membandingkan label aktual dengan prediksi model, yang terdiri atas empat komponen utama. True Positive (TP) adalah jumlah teks AI yang terdeteksi dengan benar sebagai teks AI, sedangkan True Negative (TN) adalah jumlah teks manusia yang berhasil dikenali dengan tepat sebagai teks manusia [18]. Sementara itu, False Positive (FP) menggambarkan jumlah teks manusia yang keliru diprediksi sebagai teks AI, dan False Negative (FN) menunjukan jumlah teks AI yang keliru diklasifikasikan sebagai teks manusia [19]. Hasil lengkap dari confusion matrix ditampilkan pada tabel berikut:



Gambar 5. Hasil AI vs Human Text

Hasil evaluasi menunjukkan bahwa model memiliki performa yang sangat baik dalam mengidentifikasi teks yang dihasilkan oleh AI. Hal ini ditunjukkan oleh tingginya jumlah *True Positive* (60.962) dan rendahnya jumlah *False Negative* (150) [20]. Jumlah *False Positive* yang sangat kecil (488) menunjukkan bahwa model jarang salah dalam mengklasifikasikan teks yang ditulis oleh manusia sebagai teks buatan AI. Selain itu, tingginya nilai *True Negative* (35.847) mengindikasikan bahwa model memiliki kemampuan yang sangat baik dalam mengenali dan

mengklasifikasikan teks manusia secara akurat. Hasil klasifikasi disajikan dalam bentuk tabel sebagai berikut:

Tabel 1. Hasil Klasifikasi

	Prediksi: AI	Prediksi: Human
Aktual: AI	60.962 (TP)	150 (FN)
Aktual: Human	488 (FP)	35.847 (TN)

Secara keseluruhan, *confusion matrix* menunjukkan bahwa model klasifikasi memiliki kinerja yang sangat baik dalam membedakan antara teks yang ditulis oleh manusia dan teks yang dihasilkan oleh AI, dengan tingkat kesalahan yang sangat rendah Dalam evaluasi ini, *F1-score* dipilih sebagai metrik utama karena mampu menggabungkan aspek *precision* dan *recall*, sehingga memberikan gambaran yang lebih seimbang dan menyeluruh terhadap performa model secara keseluruhan. Meskipun performa model tergolong tinggi, masih terdapat sejumlah kecil kesalahan klasifikasi, yaitu 150 kasus False Negative dan 488 kasus False Positive. Kesalahan ini kemungkinan disebabkan oleh adanya teks AI yang memiliki gaya bahasa sangat menyerupai tulisan manusia, misalnya dengan menambahkan variasi gaya tutur atau ekspresi emosional secara alami, sehingga sulit dibedakan oleh model. Sebaliknya, beberapa teks manusia yang ditulis dengan gaya formal, terstruktur, dan minim variasi gaya bahasa juga dapat menyerupai karakteristik khas teks AI, yang mengakibatkan model melakukan klasifikasi yang keliru.

Evaluasi ini digunakan untuk menilai akurasi prediksi dan mengenali kesalahan model. Setelah proses pelatihan, diperoleh hasil evaluasi sebagai berikut.

- Akurasi Model: Model mencapai akurasi sebesar 99,35%, menandakan kemampuannya yang sangat baik dalam mengklasifikasikan teks.
- Classification Report:

Tabel 2. Klasifikasi

Kelas	Precision	Recall	F1-Score	Support
Human (0.0)	0.99	1.00	0.99	61,112
AI (1.0)	1.00	0.99	0.99	36,335

Model menunjukkan keseimbangan yang sangat baik antara nilai *precision* dan *recall* pada kedua kelas, yaitu teks buatan AI dan teks manusia yang dapat diidentifikasi secara konsisten dan akurat oleh model. Temuan ini juga menunjukkan bahwa pendekatan berbasis NLP menggunakan TF-IDF dan algoritma Multinomial Naive Bayes mampu mengidentifikasi perbedaan linguistik antara teks manusia dan AI secara efektif. Dengan performa yang tinggi dan kesalahan yang minimal, model ini memiliki potensi besar untuk diterapkan dalam sistem deteksi otomatis pada berbagai bidang, seperti pendidikan, jurnalistik, dan moderasi konten digital. Di masa depan, pengembangan model yang lebih sensitif terhadap konteks dan nuansa budaya dalam bahasa dapat lebih meningkatkan akurasi serta memperkuat keandalan sistem deteksi teks.

Penelitian ini hanya menggunakan artikel dan teks berbahasa Inggris, sehingga hasil dan model yang dibangun belum mencerminkan performa atau efektivitas pada bahasa lain. Setiap bahasa memiliki karakteristik linguistik dan struktur sintaksis yang unik, sehingga pendekatan yang digunakan dalam penelitian ini belum tentu memberikan hasil yang sama jika diterapkan pada teks

berbahasa Indonesia, Mandarin, Arab, atau bahasa lainnya. Meskipun jumlah data yang digunakan dalam penelitian ini tergolong besar, distribusi antara teks buatan manusia dan teks buatan AI tidak sepenuhnya seimbang. Dataset terdiri dari 305.797 teks manusia dan 181.438 teks AI, sehingga terdapat dominasi pada salah satu kelas. Ketidakseimbangan ini berpotensi memengaruhi performa model, terutama pada metrik seperti precision dan recall yang sensitif terhadap distribusi kelas.

IV. KESIMPULAN

Penelitian ini mengkaji perbedaan linguistik antara teks yang ditulis oleh manusia dan yang dihasilkan oleh AI dengan memanfaatkan pendekatan interdisipliner yang menggabungkan teknik NLP dan perspektif kajian humaniora. Hasil analisis menunjukkan bahwa teks AI memiliki struktur yang teratur dan konsisten, tetapi cenderung datar dan kurang menampilkan ekspresi emosional maupun kedalaman konteks budaya. Sebaliknya, teks manusia menunjukkan keragaman dalam gaya bahasa, pilihan kata, dan penyampaian makna, yang lebih mencerminkan pengalaman pribadi dan nilai sosial. Pendekatan teknis melalui NLP memungkinkan identifikasi pola-pola linguistik yang membedakan keduanya secara statistik, sementara perspektif humaniora membantu memahami intensi dan nilai di balik teks, yang tidak bisa ditangkap hanya melalui angka.

Melalui model klasifikasi Multinomial Naive Bayes dan fitur TF-IDF, sistem berhasil membedakan teks AI dan manusia dengan tingkat akurasi sebesar 99,35% dan F1-score sebesar 0,9948. Confusion matrix menunjukkan jumlah True Positive yang tinggi (60.962) dan True Negative yang besar (35.847), serta jumlah kesalahan yang relatif kecil (150 False Negative dan 488 False Positive). Hal ini menunjukkan bahwa pendekatan NLP cukup andal untuk mendeteksi perbedaan linguistik antara dua jenis teks tersebut. Namun, tetap terdapat tantangan ketika teks AI meniru gaya manusia secara natural, atau ketika teks manusia ditulis dengan struktur yang menyerupai pola bahasa mesin.

Berdasarkan hasil tersebut, penelitian ini merekomendasikan agar pendekatan ke depan menggunakan model semantik lanjutan seperti BERT untuk menangkap konteks secara lebih akurat. Selain itu, perluasan dataset dengan variasi domain dan pengujian lintas bahasa akan meningkatkan generalisasi model. Evaluasi terhadap dampak preprocessing juga penting, mengingat proses seperti stemming atau stopword removal dapat menghapus ciri khas penting dalam teks manusia. Terakhir, pengembangan sistem deteksi otomatis berbasis AI perlu mempertimbangkan transparansi dan etika, khususnya untuk aplikasi di bidang pendidikan, media, dan pemantauan konten digital, agar teknologi tetap berpihak pada orisinalitas dan nilai kemanusiaan.

REFERENSI

[1] Agusman, M. D. (2025). Perlindungan hak cipta berbasis NFT dan smart contract dalam menanggapi isu pencurian suatu karya digital. *Jurnal Hukum*, 4(2), 387–394.

[2] Cunliffe, D., Vlachidis, A., Williams, D., & Tudhope, D. (2022). Natural language processing for under-resourced languages: Developing a Welsh natural language toolkit. *Computer Speech & Language*, 72, 101311. https://doi.org/10.1016/j.csl.2021.101311

[3] Zhou, C., et al. (2023). A comprehensive survey on pretrained foundation models: A history from BERT to ChatGPT. *Artificial Intelligence Review*, 1–99. https://doi.org/10.1007/s13042-024-02443-6

[4] Wang, H., Li, J., & Li, Z. (2024). AI-generated text detection and classification based on BERT deep learning algorithm. *Theoretical and Natural Science*, 39(1), 187–192. https://doi.org/10.54254/2753-8818/39/20240625

[5] Noor, N., & Prova, I. (n.d.). Detecting AI generated text based on NLP and machine learning approaches. *Preprint*.

[6] Sains, F., Teknologi, D. A. N., Ar-Raniry, U. I. N., & Aceh, B. (2024). Perbandingan metode Support Vector Machine dan Naive Bayes terhadap penggunaan Artificial Intelligence dalam pembuatan skripsi pada media sosial X.

[7] Liu, J., Nie, Y., & Chua, B. L. (2024). Generative AI in assessment: AI detectors and implications for practice. *Preprint*, 1–20.

[8] Sulartopo, S., Kholifah, S., Danang, D., & Santoso, J. T. (2023). Transformasi proyek melalui keajaiban kecerdasan buatan: Mengeksplorasi potensi AI dalam project management. *Jurnal Publikasi Ilmu Manajemen*, 2(2), 363–392.

[9] Picciotto, H., & Pemantle, R. (2024). Learning tools. In *There Is No One Way to Teach Math* (Vol. 4, No. 6, pp. 77–98). https://doi.org/10.4324/9781003473855-8

[10] Georgiou, G. P. (2023). Differentiating between human-written and AI-generated texts using linguistic features automatically extracted from an online computational tool. *Manuscript*, 1–12.

[11] Sandler, M., Choung, H., Ross, A., & David, P. (2024). A linguistic comparison between human and ChatGPT-generated conversations. In *Proceedings of the Conference on Computational Linguistics*, 1–15. https://doi.org/10.1007/978-981-97-8702-9_25

[12] Muñoz-Ortiz, A., Gómez-Rodríguez, C., & Vilares, D. (2024). Contrasting linguistic patterns in human and LLM-generated news text. *Artificial Intelligence Review*, 57(9), 1–28. https://doi.org/10.1007/s10462-024-10903-2

[13] Sitanggang, A., Umaidah, Y., Adam, R. I. (2024). Analisis sentimen masyarakat terhadap program makan siang gratis pada media sosial X menggunakan algoritma Naïve Bayes. *Jurnal Informatika dan Teknik Elektro Terapan*, 12(3). https://doi.org/10.23960/jitet.v12i3.4902

[14] Fariello, S., Fenza, G., Forte, F., Gallo, M., & Marotta, M. (2024). Distinguishing human from machine: A review of advances and challenges in AI-generated text detection. *International Journal of Interactive Multimedia and Artificial Intelligence*, January. https://doi.org/10.9781/ijimai.2024.12.002

[15] Gunawan, A., Altiarika, E., Pratama, S., & Pratama, Y. B. (2024). Pengembangan aplikasi asisten virtual menggunakan machine learning berbasis mobile untuk meningkatkan pelayanan kampus di Muhammadiyah Bangka Belitung. *Jurnal Teknologi Informasi*, 2(2), 45–51.

[16] Prismala, D., & Nuryana, I. K. D. (2024). Analisis opinion mining pada topik ChatGPT di aplikasi X dengan pendekatan algoritma SVM berbasis lexicon. *Jurnal Teknologi Informasi*, 6, 479–492.

[17] Universitas Muhammadiyah Aceh & Universitas Bina Nusantara. (2024). Penggunaan algoritma Support Vector Machine (SVM) untuk deteksi penipuan pada transaksi online. *Jurnal Teknologi dan Sistem Komputer*, 13, 1627–1632.

[18] Widaad, N., Anggraini, D., & Faculty of Engineering, Universitas Gunadarma. (2024). Sentiment analysis of ChatGPT app user reviews using SVM and CNN. *Jurnal Ilmu Komputer*, 5(6), 1687–1700.

[19] Widiawati, F., Kurniawan, R., & Suprapti, T. (2024). Klasifikasi data tingkat kualitas udara di Tangerang Selatan menggunakan algoritma Naive Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3739–3745. https://doi.org/10.36040/jati.v7i6.8261

[20] Putri Kumala Sari, R. R. S. (2024). Komparasi algoritma Support Vector Machine dan Random Forest. *Jurnal Teknologi Informasi*, 7(1), 31–39.