

KLASIFIKASI PROMPT PADA AI GENERATE IMAGE DENGAN METODE GAUSSIAN NAÏVE BAYES DAN DECISION TREE CLASSIFIER

Putri Paramitha ^{1*}, Agung Riyadi ^{2**}

* Teknik Informatika, Politeknik Negeri Batam

** Teknik Informatika, Politeknik Negeri Batam

putriparamithaa@gmail.com ¹, agung@polibatam.ac.id ²

Article Info

Article history:

Received ...

Revised ...

Accepted ...

Keyword:

Pilih maksimum lima kata kunci atau frase yang diurutkan menurut abjad, dan dipisahkan dengan koma. *Keyword1, Keyword2, Keyword3.*

ABSTRACT

Perkembangan teknologi Kecerdasan Buatan (AI), khususnya dalam domain AI Generate Image, telah memungkinkan pengguna untuk menghasilkan visual dari teks dengan cepat dan efisien. Namun, tantangan dalam klasifikasi gambar yang akurat masih menjadi perhatian utama. Penelitian ini bertujuan untuk mengeksplorasi dan membandingkan dua metode klasifikasi, yaitu Gaussian Naïve Bayes dan Decision Tree Classifier, dalam mengelompokkan gambar ke dalam tiga kategori: Background/Texture, Landscape, dan Arts. Proses pengujian dilakukan untuk mengevaluasi akurasi serta efektivitas kedua metode dalam mengklasifikasikan data gambar yang tersedia. Hasil analisis menunjukkan bahwa Gaussian Naïve Bayes mencapai akurasi sebesar 61.74%, sementara Decision Tree Classifier menunjukkan performa yang lebih unggul dengan akurasi sebesar 99.16%. Meskipun kedua metode diuji berdasarkan standar akurasi minimum 90%, hanya Decision Tree Classifier yang berhasil melampaui batas tersebut, menunjukkan keunggulan dalam hal akurasi dan kinerja keseluruhan. Penelitian ini diharapkan dapat memberikan kontribusi yang berarti dalam meningkatkan kinerja sistem AI Generate Image, serta dalam pengelolaan dan penyajian gambar yang lebih terstruktur dan sesuai dengan kategori yang diinginkan.

I. PENDAHULUAN

Revolusi Industri 4.0 telah membawa perubahan besar dalam teknologi informasi, termasuk perkembangan pesat dalam Artificial Intelligence (AI). Salah satu inovasi terkini dalam AI adalah kemampuan untuk menghasilkan gambar dari teks melalui aplikasi AI Generate Image. Teknologi ini memungkinkan pengguna untuk membuat visual yang relevan dengan deskripsi teks dalam waktu singkat, yang membuka peluang baru dalam pengelolaan dan pemanfaatan data visual.

Namun, meskipun teknologi ini menawarkan kemudahan, klasifikasi gambar yang dihasilkan masih menghadapi tantangan dalam hal akurasi dan efisiensi. Klasifikasi yang tepat sangat penting untuk memastikan bahwa gambar dikelompokkan dengan benar ke dalam kategori yang diinginkan, seperti Background/Texture, Landscape, dan Arts. Kesalahan dalam pengelompokan dapat mengakibatkan kesulitan dalam mencari dan menggunakan gambar sesuai kebutuhan.

Penelitian ini bertujuan untuk mengevaluasi dan membandingkan dua metode klasifikasi gambar

menggunakan machine learning, yaitu Gaussian Naïve Bayes dan Decision Tree Classifier. Keduanya dipilih karena kemampuannya dalam menangani masalah klasifikasi dan perbedaan pendekatan yang mereka gunakan dalam pengelompokan data. Penelitian ini akan menilai efektivitas masing-masing metode dalam mengelompokkan gambar yang dihasilkan oleh AI dan membandingkan hasil akurasi dari kedua metode tersebut.

Dengan menganalisis dan membandingkan kedua metode klasifikasi, penelitian ini diharapkan dapat memberikan wawasan tentang metode yang paling efektif untuk meningkatkan akurasi dan efisiensi dalam sistem AI Generate Image. Hasil penelitian ini diharapkan tidak hanya memperbaiki kinerja klasifikasi gambar tetapi juga meningkatkan pengalaman pengguna dalam mengelola dan mencari gambar sesuai kategori yang diinginkan.

II. LANDASAN TEORI

A Tinjauan Pustaka

Berbagai penelitian telah dilakukan untuk mengeksplorasi metode klasifikasi data menggunakan algoritma machine learning dalam konteks yang berbeda. Misalnya, penelitian oleh Ramadhan Rakhmat Sani dan rekan-rekannya yang berjudul "Analisis Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Hoax pada Berita Online Indonesia" menguji algoritma Naive Bayes dan Support Vector Machine (SVM) untuk mengklasifikasi berita hoax. Dalam penelitian ini, SVM menunjukkan kinerja yang lebih baik dibandingkan Naive Bayes, terutama pada isu kesehatan, dengan menggunakan TF-IDF untuk meningkatkan precision dan recall.

Penelitian lain oleh Nabila Bianca Putri dan Arie Wahyu Wijayanto membandingkan beberapa algoritma klasifikasi seperti Naive Bayes, Random Forest, Decision Tree, dan SVM dalam konteks klasifikasi website phishing. Hasilnya menunjukkan bahwa Random Forest memiliki akurasi tertinggi, mencapai 95,61%, sementara Naive Bayes memiliki akurasi terendah sebesar 82,31%. Selanjutnya, penelitian yang dilakukan oleh Ahmad Firdaus dan koleganya menggunakan algoritma Gaussian Naive Bayes untuk mengklasifikasikan kasus Covid-19 di Kabupaten Pamekasan. Algoritma ini terbukti sangat efektif dengan akurasi 96,88%, precision 0,97, recall 1,00, dan f1-score 0,98.

Selain itu, Agus Fajar Riany dan Gusmelia Testiana melakukan penelitian terkait klasifikasi penyakit stroke menggunakan algoritma Naive Bayes, yang mencapai akurasi sebesar 92,48%, membuktikan bahwa algoritma ini cocok untuk klasifikasi data medis. Secara keseluruhan, penelitian-penelitian ini menunjukkan bahwa algoritma machine learning, khususnya Naive Bayes dan SVM, sangat berguna dalam klasifikasi data di berbagai domain, meskipun tingkat keberhasilannya sangat bergantung pada karakteristik data yang dianalisis. Berikut adalah detail dari penelitian yang terkait, dapat dilihat pada tabel 1.

Tabel 1 Perbandingan Klasifikasi

N o	Judul	Algoritma	Peneliti
1	Analisis Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Hoax pada	<i>Multinomial Naive Bayes Classifier</i> dan <i>Support Vector Machine</i> .	Ramadhan Rakhmat Sani, Yunita Ayu Pratiwi, Sri Winarno, Erika Devi

	Berita Online Indonesia.		Udayanti, dan Farrikh Al Zami.
2	Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing.	<i>Naive Bayes, Random Forest, Decision Trees, Support Vector Machine.</i>	Nabila Bianca Putri dan Arie Wahyu Wijayanto
3	Klasifikasi Kasus Covid-19 Menggunakan Model Naive Bayes Classifier.	<i>Naive bayes (Gaussian Naive Bayes).</i>	Ahmad Firdaus dan Miftahul Walid, Anwari.
4	Penerapan Data Mining untuk Klasifikasi Penyakit Stroke Menggunakan Algoritma Naive Bayes.	<i>Naive Bayes</i>	Agus Fajar Riany, Gusmelia Testiana.

B. Dasar Teori

1. AI (Artificial Intelligence)

Artificial Intelligence (AI) adalah teknologi yang memungkinkan mesin melakukan tugas-tugas yang biasanya membutuhkan kecerdasan manusia, seperti memahami bahasa, mengenali input audio dan visual, membuat keputusan, serta menyelesaikan masalah. AI menggunakan algoritma pembelajaran mesin dan model statistik untuk belajar dari data, serta meningkatkan kinerjanya secara bertahap seiring waktu. Teknologi ini dapat memproses data dalam jumlah besar, menemukan pola, dan menghasilkan prediksi atau keputusan berdasarkan analisis tersebut.

Menurut McCarthy (1956), AI adalah ilmu dan teknik dalam menciptakan mesin cerdas, terutama melalui pengembangan program dan aplikasi komputer yang dapat meniru kecerdasan manusia. Beberapa penerapan umum AI termasuk pemrosesan bahasa alami, pengenalan gambar, pemberian rekomendasi, pendeteksi penipuan, dan teknologi kendaraan otonom.

2. AI Generative Image

AI Generative Image adalah teknologi yang mampu menghasilkan gambar berdasarkan perintah berupa teks, menggunakan model pembelajaran mendalam yang telah dilatih dengan data skala besar. Teknologi ini memanfaatkan

model bahasa visual multimodal untuk menciptakan representasi gambar digital yang realistis dan kompleks, termasuk lanskap, objek, dan wajah. Tujuan utama AI Generative Image adalah menghasilkan gambar baru yang belum pernah ada, dengan kemampuan mengekstrak pola dari berbagai input seperti teks, visual, atau audio. Aplikasi AI ini luas, termasuk pembuatan gambar 3D, konten visual untuk blog, hingga media periklanan, yang menjadikannya alat penting dalam industri kreatif.

3. Data Mining

Data Mining merupakan proses semi-otomatis yang memanfaatkan berbagai teknik dari disiplin ilmu seperti statistika, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstraksi serta mengidentifikasi pengetahuan tersembunyi yang bermanfaat dari kumpulan data berukuran besar. Menurut Turban et al. (2005), proses ini memungkinkan penemuan informasi yang signifikan dan berguna melalui analisis sistematis. Data mining juga menjadi bagian integral dari proses Knowledge Discovery in Databases (KDD), yang mencakup beberapa tahapan kunci, yakni pemilihan data, pra-pemrosesan, transformasi, proses data mining itu sendiri, dan evaluasi hasil (Maimon & Last, 2000). KDD secara umum dikenal sebagai proses yang berfungsi untuk mengelola dan mengekstraksi pengetahuan dari basis data.

Dalam data mining, terdapat dua pendekatan utama, yaitu Supervised Methods dan Unsupervised Methods. Pada Supervised Methods, variabel target sudah ditentukan sebelumnya, sehingga algoritma yang digunakan dapat belajar dari hubungan antara variabel target dan variabel prediktor yang ada. Sebaliknya, pada Unsupervised Methods, tidak ada variabel target yang ditentukan, sehingga pendekatan ini bertujuan menemukan pola atau struktur yang tersembunyi dalam data tanpa pengawasan.

4. Klasifikasi

Klasifikasi merupakan proses untuk menemukan sekumpulan fungsi yang dapat menjelaskan dan membedakan suatu konsep atau kelas dalam data, dengan tujuan mengklasifikasikan data baru. Proses ini memerlukan manipulasi data yang telah melalui proses pengelompokan sebelumnya, di mana hasil akhirnya adalah aturan-aturan yang digunakan untuk klasifikasi lebih lanjut. Berbeda dengan clustering, yang mengelompokkan data tanpa adanya variabel target, klasifikasi dilakukan berdasarkan variabel tertentu yang telah ditentukan sebelumnya.

Proses klasifikasi biasanya terdiri dari dua tahap utama. Tahap pertama adalah learning, di mana algoritma klasifikasi menganalisis data pelatihan (training data) untuk membangun model prediktif. Tahap kedua adalah klasifikasi, di mana model yang dibangun dari data pelatihan digunakan

untuk memprediksi hasil pada data uji (testing data). Pada akhirnya, klasifikasi bertujuan untuk memprediksi ketepatan model dalam mengklasifikasikan data baru berdasarkan aturan yang dihasilkan dari data pelatihan tersebut.

5. Bag of Word

Bag of Words (BoW) merupakan suatu teknik paling sederhana dalam representasi teks untuk keperluan klasifikasi dalam data mining. Teknik ini mengubah teks menjadi vektor numerik berdasarkan frekuensi kemunculan kata dalam dokumen tanpa memperhatikan urutan atau konteks kata. Menurut Manning et al. (2008), BoW mengabaikan struktur kalimat dan hanya fokus pada seberapa sering kata muncul, sehingga menghasilkan vektor berdimensi tinggi yang dapat digunakan oleh algoritma pembelajaran mesin, seperti Naive Bayes atau Decision Tree.

Proses BoW melibatkan langkah tokenisasi teks menjadi kata-kata, filtering untuk menghilangkan kata-kata yang tidak relevan (stopwords), dan akhirnya, menghitung frekuensi setiap kata menggunakan metode CountVectorizer. Vektor ini kemudian digunakan sebagai input untuk algoritma pembelajaran mesin dalam proses klasifikasi.

Meskipun BoW mudah diimplementasikan dan efektif dalam berbagai aplikasi seperti klasifikasi teks, teknik ini memiliki keterbatasan, seperti tidak mempertimbangkan urutan kata dan konteks makna, yang bisa mempengaruhi hasil analisis jika konteks sangat penting.

6. Gaussian Naive Bayes

Gaussian Naive Bayes adalah algoritma klasifikasi yang sering digunakan dalam situasi di mana fitur-fitur data bersifat kontinu dan diasumsikan mengikuti distribusi normal. Algoritma ini bekerja dengan menghitung probabilitas kelas-kelas yang mungkin berdasarkan fitur input yang diberikan, dan kemudian memilih kelas dengan probabilitas tertinggi sebagai prediksi akhir. Meskipun pendekatannya tergolong sederhana, terutama karena asumsi independensi antar fitur yang disebut "naif," Gaussian Naive Bayes terbukti efektif dalam berbagai aplikasi, termasuk klasifikasi teks dan penyaringan spam.

Model Gaussian Naive Bayes memanfaatkan dua parameter utama, yaitu rata-rata (mean) dan variansi (variance) dari distribusi Gaussian. Dalam konteks ini, distribusi Gaussian atau normal digambarkan dengan persamaan berikut:

$$P(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Distribusi ini berasumsi bahwa setiap fitur dalam data memberikan kontribusi independen terhadap prediksi target. Proses klasifikasi dilakukan dengan menggabungkan

prediksi dari seluruh fitur untuk menghasilkan probabilitas akhir dari variabel target yang dikategorikan ke dalam kelas-kelas tertentu. Kelas dengan probabilitas tertinggi menjadi hasil klasifikasi akhir.

7. Decision Tree classifier

Decision Tree Classifier adalah algoritma yang menggunakan struktur pohon untuk membuat keputusan klasifikasi. Setiap simpul dalam pohon mewakili atribut data, sementara cabang menunjukkan keputusan berdasarkan hasil pengujian atribut tersebut. Proses dimulai dari simpul akar yang membagi dataset secara bertahap hingga mencapai simpul daun yang menunjukkan kelas akhir.

Algoritma ini menggunakan metrik seperti Gini Impurity atau Information Gain untuk menentukan pemisahan terbaik di setiap simpul. Kelebihan utama Decision Tree adalah kemampuannya untuk menangani variabel kategorik dan kontinu, serta interpretasinya yang mudah melalui visualisasi pohon keputusan.

III. METODE

A. Pengumpulan Data

Data yang digunakan dalam penelitian ini diambil dari website AI Generate Image Prompts. Data tersebut mencakup tiga kategori, yaitu background/texture, landscape, dan art, dengan total 7040 entri.

B. Preprocessing Data

Tahap preprocessing dilakukan untuk mempersiapkan data sebelum analisis. Proses ini mencakup beberapa langkah:

1. Pembersihan Data

Langkah pertama dalam mempersiapkan dataset adalah memastikan data yang akan digunakan bersih dan berkualitas. Pada tahap ini, data yang tidak lengkap, duplikat, atau tidak relevan dihapus dari dataset. Data yang tidak lengkap biasanya muncul dalam bentuk baris yang memiliki nilai kosong atau atribut yang hilang, yang dapat mempengaruhi hasil klasifikasi jika tidak ditangani dengan benar. Oleh karena itu, entri semacam ini dihapus atau diperbaiki dengan nilai yang sesuai. Selain itu, baris duplikat juga dihapus karena dapat menyebabkan model machine learning mengalami overfitting. Data yang tidak relevan, seperti entri yang tidak sesuai dengan kategori penelitian yang diinginkan (background/texture, landscape, art), juga dihilangkan, sehingga dataset lebih sesuai dengan tujuan klasifikasi yang akan dilakukan.

Tahapan ini dapat dilakukan secara manual yaitu dengan menggunakan rumus penghapusan data duplikat pada Excel, dikarenakan data awalnya masih berupa

tabel pada Excel yang mana diberi angka 0 sebagai tanda untuk data Art, angka 1 untuk data Background, dan angka 2 sebagai tanda untuk data Landscape. Langkah ini juga dapat dilakukan secara otomatis dan lebih cepat yaitu dengan memanfaatkan bahasa pemrograman Python.

2. Transformasi Data

Setelah data dibersihkan, langkah selanjutnya adalah mengubah format data dari Excel ke TSV (Tab-Separated Values). Format TSV dipilih karena lebih mudah diproses oleh algoritma machine learning. Selain mengubah format file, tahap ini juga melibatkan penyaringan data untuk memastikan hanya data yang relevan yang digunakan dalam penelitian. Dari data awal yang berjumlah 7040 entri, setelah pembersihan dan transformasi, data yang layak digunakan berkurang menjadi 4768 entri. Data berkurang disebabkan oleh penghapusan entri yang tidak valid atau tidak sesuai selama proses pembersihan. Dataset yang sudah ditransformasi selanjutnya akan digunakan dalam proses klasifikasi dan analisis lebih lanjut.

3. Penggunaan Tools

Seluruh proses mulai dari pembersihan data hingga analisis dilakukan menggunakan **Visual Studio Code**. Platform ini dipilih karena fleksibilitasnya dalam mendukung berbagai bahasa pemrograman, terutama Python, yang digunakan dalam pembersihan dan klasifikasi data. Dengan berbagai ekstensi yang mendukung pustaka seperti Pandas, Numpy, dan Scikit-learn, Visual Studio Code menyediakan lingkungan yang terpadu dan efisien bagi peneliti untuk menulis, menjalankan, dan memverifikasi kode. Kemudahan dalam menjalankan skrip analisis data langsung di satu platform membuat proses penelitian menjadi lebih cepat dan efisien.

C. Representasi Data dengan Bag of Word

Bag of Words (BoW) adalah metode yang umum digunakan dalam pengolahan bahasa alami (NLP) untuk mengubah teks menjadi bentuk numerik yang bisa dipahami oleh algoritma machine learning. Dalam BoW, setiap dokumen atau teks diwakili sebagai daftar kata-kata unik yang ada di dalamnya, tanpa memperhatikan urutan kata. Setiap kata dianggap sebagai fitur, dan yang dihitung adalah frekuensi kemunculan kata-kata tersebut dalam dokumen.

Tahap pertama dalam BoW adalah tokenisasi, yaitu memecah teks menjadi kata-kata individu. Proses ini bertujuan untuk memisahkan satu persatu kata dalam sebuah teks, contohnya pada kalimat “comic panels, invoking the dark arts” menjadi (“comic”, “panels”, “invoking”, “the”, “dark”, “arts”). Setelah itu, dilakukan pembuangan

stopwords, di mana kata-kata umum seperti “the”, “is”, atau “and” yang tidak memiliki makna signifikan dalam klasifikasi dihapus. Tahap berikutnya adalah stemming, di mana kata-kata tersebut dikembalikan ke bentuk dasarnya (misalnya, “Baked” menjadi “Bake”), untuk mengurangi variasi dan menyederhanakan analisis.

Hasil dari proses ini adalah sebuah matriks, di mana setiap baris mewakili satu dokumen, dan setiap kolom mewakili kata unik yang ada di korpus. Nilai di dalam matriks menunjukkan seberapa sering kata tersebut muncul dalam dokumen. Matriks ini memungkinkan algoritma klasifikasi seperti Gaussian Naive Bayes atau Decision Tree untuk memproses data teks dan melakukan analisis lebih lanjut.

Dalam penelitian ini, CountVectorizer digunakan untuk membangun representasi BoW, dengan membatasi jumlah fitur maksimal hingga 7332. BoW membantu dalam menganalisis teks deskripsi gambar dari dataset yang telah dikumpulkan dan memudahkan algoritma dalam mengelompokkan gambar berdasarkan kategori yang relevan.

D. Teknik Klasifikasi

Analisis data dalam penelitian ini dilakukan menggunakan dua pendekatan klasifikasi utama, yaitu Gaussian Naive Bayes dan Decision Tree Classifier. Kedua metode ini digunakan untuk menentukan kategori dari data berdasarkan fitur yang diperoleh dari teks deskripsi gambar yang akan dijabarkan sebagai berikut.

1. Gaussian Naive Bayes

Algoritma ini merupakan varian dari Naive Bayes yang menggunakan teorema Bayes dengan asumsi bahwa fitur yang diukur mengikuti distribusi Gaussian (normal). Metode ini efektif untuk dataset dengan variabel kontinu dan menghasilkan klasifikasi yang cepat. Dengan menggunakan probabilitas, Gaussian Naive Bayes memperkirakan kemungkinan suatu data termasuk ke dalam kategori tertentu, berdasarkan fitur-fitur yang ada.

2. Decision Tree Classifier

Metode ini memanfaatkan struktur pohon keputusan, di mana setiap cabang dalam pohon mewakili aturan dan setiap simpul terminal mewakili hasil akhir klasifikasi. Decision Tree Classifier bekerja dengan cara membagi dataset menjadi subset yang lebih kecil berdasarkan fitur-fitur yang paling relevan, sehingga memudahkan dalam pengambilan keputusan. Struktur yang intuitif membuatnya mudah untuk diinterpretasikan, dan metode ini sering digunakan untuk berbagai aplikasi klasifikasi.

E. Pelatihan dan Pengujian Model (Training dan Testing)

Setelah menerapkan algoritma klasifikasi, data yang sudah diolah (representasi numerik dari teks) dibagi menjadi dua set: training set untuk melatih model dan testing set untuk menguji performa model. Pada tahap ini, model dilatih menggunakan data latih dan kemudian diuji menggunakan data uji untuk melihat seberapa baik model dapat memprediksi kelas data yang tidak pernah dilihat sebelumnya. Dimana data yang digunakan untuk testing sebanyak 20%, sedangkan sisanya untuk training sebanyak 80% dari keseluruhan data.

F. Evaluasi Model (Model Evaluation)

Pada tahap ini, performa model dievaluasi dengan menggunakan beberapa metrik seperti akurasi, precision, recall, F1-score, dan confusion matrix. Evaluasi ini penting untuk mengetahui seberapa efektif model dalam mengklasifikasikan data.

1. Akurasi: Mengukur persentase prediksi yang benar.
2. Precision dan Recall: Mengukur seberapa baik model menangani kelas minoritas atau kejadian langka.
3. F1-score: Kombinasi precision dan recall untuk memberikan gambaran seimbang mengenai performa model.
4. Confusion Matrix: Matriks yang menunjukkan jumlah prediksi yang benar dan salah untuk masing-masing kelas.

IV. HASIL DAN PEMBAHASAN

A. Hasil Klasifikasi

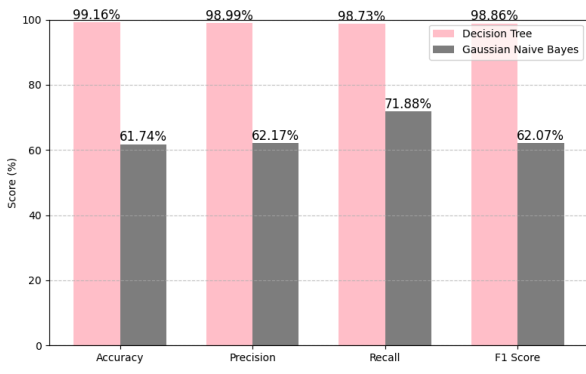
Penelitian ini mengevaluasi dua metode klasifikasi, yaitu Gaussian Naive Bayes dan Decision Tree Classifier, untuk mengklasifikasikan gambar ke dalam tiga kategori: Background/Texture, Landscape, dan Arts. Hasil dari evaluasi kedua model ditampilkan dalam Tabel 1 berikut ini:

Metode Klasifikasi	Akurasi (%)	Precision (%)	Recall (%)	F1 Score (%)
Gaussian Naive Bayes	61.74	62.17	71.88	62.07

Decision Classifier	Tree	99.16	98.99	98.73	98.86
---------------------	------	-------	-------	-------	-------

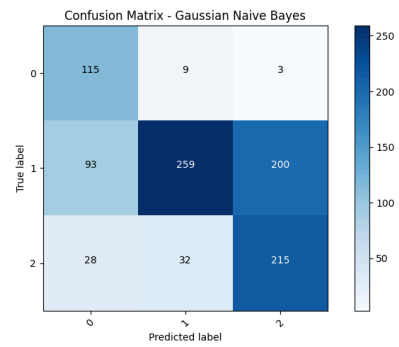
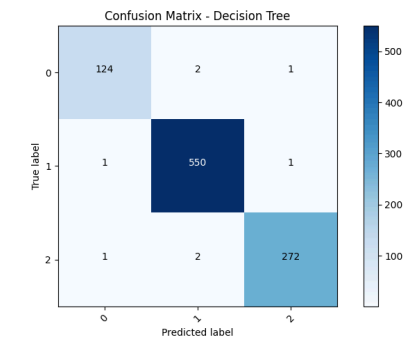
Tabel 1 menunjukkan bahwa Decision Tree Classifier memiliki performa yang jauh lebih baik dibandingkan dengan Gaussian Naive Bayes dalam semua metrik evaluasi.

Selanjutnya, Gambar 1 di bawah ini menunjukkan grafik perbandingan matrik evaluasi untuk kedua metode klasifikasi. Grafik ini memberikan gambaran yang jelas mengenai perbandingan performa antara Gaussian Naive Bayes dan Decision Tree Classifier.



B. Confusion Matrix

Confusion matrix untuk masing-masing metode klasifikasi juga dihitung untuk memberikan gambaran yang lebih jelas mengenai kinerja model. Berikut adalah Gambaran confusion matrix untuk kedua metode klasifikasi.



C. Pembahasan Hasil

Hasil klasifikasi menunjukkan bahwa Decision Tree Classifier memiliki performa yang lebih baik dengan akurasi mencapai 99.16%, sedangkan Gaussian Naive Bayes hanya mencapai 61.74%. Perbedaan ini dapat dipahami dari karakteristik masing-masing algoritma. Gaussian Naive Bayes mengasumsikan fitur-fitur bersifat independen dan mengikuti distribusi normal, yang mungkin tidak berlaku dalam dataset ini.

Sebaliknya, Decision Tree Classifier lebih fleksibel, menggunakan pendekatan berbasis aturan yang dapat menangkap pola yang lebih kompleks. Keuntungan dari Decision Tree adalah kemampuannya untuk memberikan interpretasi yang jelas tentang keputusan yang diambil, meskipun rentan terhadap overfitting jika model terlalu kompleks.

Secara keseluruhan, Decision Tree Classifier terbukti lebih efektif dalam mengidentifikasi pola data, menjadikannya pilihan yang lebih baik untuk analisis yang memerlukan akurasi tinggi dan transparansi dalam proses pengambilan keputusan.

DAFTAR PUSTAKA

- [1] Ramadhan Rakhmat Sani, Pratiwi, Y. A., Winarno, S., Udayanti, E. D., & Al Zami, F. (2022). Analisis perbandingan algoritma Naive Bayes Classifier dan Support Vector Machine untuk klasifikasi hoax pada berita online Indonesia. *Jurnal Masyarakat Informatika*, 13(2), 2086-4930.
- [2] Ahmad Firdaus, Walid, M., & Anwari. (2022). Klasifikasi kasus Covid-19 menggunakan model Naive Bayes Classifier. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 583-588.
- [3] Nabila Bianca Putri & Wijayanto, A. W. (2022). Analisis komparasi algoritma klasifikasi data mining dalam klasifikasi website phishing. *Jurnal Sistem Komputer*, 11(1), 59-66.
- [4] Agus Fajar Riany & Testiana, G. (2023). Penerapan data mining untuk klasifikasi penyakit stroke menggunakan algoritma Naive Bayes. *Jurnal SAINTEKOM (Sains, Teknologi, Komputer, dan Manajemen)*, 13(1), 42-54.
- [5] Arindi, A. S., & Mirza, A. H. (2023). Model klasifikasi kinerja pegawai dengan penerapan machine learning

- menggunakan tools Python. *Jurnal Instek (Informatika Sains dan Teknologi)*, 8(1).
- [6] Simanjuntak, A. Y., Salomo Simatupang, I. S., & Anita. (2022). Implementasi data mining menggunakan metode Naive Bayes Classifier untuk data kenaikan pangkat Dinas Ketenagakerjaan Kota Medan. *Journal of Science and Social Research*, 1, 85-91.
- [7] Saputra, B., Anwar, S., Tohidi, E., Susana, H., & Pratama, D. (2023). Penerapan algoritma Naive Bayes klasifikasi harga ponsel. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6).
- [8] RevoU. (2024). Algoritma Naive Bayes. [Online]. Diakses dari <https://revou.co/kosakata/algoritma-naive-bayes>
- [9] GeeksforGeeks. (2024). Gaussian Naive Bayes. [Online]. Diakses dari <https://www.geeksforgeeks.org/gaussian-naive-bayes/>
- [10] Scikit-learn. (2024). Naive Bayes. [Online]. Diakses dari https://scikit-learn.org/stable/modules/naive_bayes.html
- [11] Analytics Vidhya. (2024). Image classification using machine learning. [Online]. Diakses dari <https://www.analyticsvidhya.com/blog/2022/01/image-classification-using-machine-learning/>
- [12] Acer. (2024). Mengenal klasifikasi teknologi AI, contoh, dan definisi. [Online]. Diakses dari <https://www.acerid.com/bisnis/klasifikasi-teknologi-ai-dan-cointohnya>
- [13] Amrullah, N. H. (2024). Metode klasifikasi dalam data mining. Diakses dari <https://www.kompasiana.com/najahamrullah1322/62b008fe38350056ba000382/metode-klasifikasi-dalam-data-mining>
- [14] Kuncoro, A. A. (2024). Seputar data mining. Diakses dari <https://teknik-informatika-s1.stekom.ac.id/informasi/baca/Seputar-Data-Mining/b219c7efbf18abc39cf426de906e6f9fd7dab267>