

## Analisis Kinerja Algoritma *Naive Bayes* dalam Klasifikasi Data Kategorikal Prediksi Keputusan Bermain Tenis Berdasarkan Cuaca

<sup>1</sup>Feriandri Lesmana, <sup>2</sup>Athila Defian Rizkimu, <sup>3</sup>Muhamad Ridwan Nurrulloh, <sup>4</sup>Maulana Farras Fathurrahman,  
<sup>5</sup>Abdul Habib Hasibuan, <sup>6</sup>Maulana Fansyuri.

<sup>1</sup>Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

<sup>2</sup>Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

<sup>3</sup>Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

<sup>4</sup>Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

<sup>5</sup>Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

<sup>6</sup>Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

[<sup>1</sup>ferilesmana1897@gmail.com](mailto:ferilesmana1897@gmail.com), [<sup>2</sup>athila.defian@gmail.com](mailto:athila.defian@gmail.com), [<sup>3</sup>mr7018112@gmail.com](mailto:mr7018112@gmail.com),  
[<sup>4</sup>maulanafarrasfathurrahman@gmail.com](mailto:maulanafarrasfathurrahman@gmail.com), [<sup>5</sup>ucokdul82@gmail.com](mailto:ucokdul82@gmail.com), [<sup>6</sup>dosen02359@unpam.ac.id](mailto:dosen02359@unpam.ac.id)

### ABSTRACT

*Decision-making based on weather factors is often subjective and inconsistent. This research applies data mining classification methods to build an objective predictive model regarding the decision to play tennis based on weather conditions. The objective of this study is to analyze the performance of the Naive Bayes algorithm in predicting this decision. The methodology involves applying the Naive Bayes algorithm to the classic "Play Tennis" dataset, which consists of 14 instances with four categorical predictor attributes: outlook, temperature, humidity, and wind. The modeling and evaluation process was conducted visually using the Altair AI Studio (RapidMiner) platform, employing the cross-validation technique to test model stability. The test results show an average model accuracy of 57.14%. A deeper analysis of the confusion matrix reveals that the model has a strong bias towards predicting the 'Yes' class, yet is very weak in identifying the 'No' class (20.00% recall). Specifically, the model exhibits a high number of False Positive errors, where 4 out of 5 'No' cases were misclassified. In conclusion, the Naive Bayes model in its current configuration is not yet fully reliable for practical application due to its biased performance. This study recommends further optimization, such as applying data balancing techniques or using more complex alternative algorithms, to significantly improve predictive performance.*

**Keywords:** Classification, Naive Bayes, Data Mining, RapidMiner, Weather Prediction.

### ABSTRAK

Pengambilan keputusan berdasarkan faktor cuaca seringkali bersifat subjektif dan inkonsisten. Penelitian ini menerapkan metode klasifikasi data mining untuk membangun model prediksi objektif terkait keputusan bermain tenis berdasarkan kondisi cuaca. Tujuan dari penelitian ini adalah untuk menganalisis kinerja algoritma *Naive Bayes* dalam memprediksi keputusan tersebut. Metodologi yang digunakan adalah penerapan algoritma *Naive Bayes* pada dataset klasik "Play Tennis", yang terdiri dari 14 instance dengan empat atribut prediktor kategorikal: outlook, temperature, humidity, dan wind. Proses pemodelan dan evaluasi dilakukan secara visual menggunakan platform Altair AI Studio (*RapidMiner*), dengan teknik validasi silang (cross-validation) untuk menguji kestabilan model. Hasil pengujian menunjukkan tingkat akurasi rata-rata model sebesar 57.14%. Analisis lebih dalam pada confusion matrix mengungkapkan bahwa model memiliki bias yang kuat dalam memprediksi kelas 'Yes', namun sangat lemah dalam mengidentifikasi kelas 'No' (recall 20.00%). Secara spesifik, model ini menunjukkan jumlah kesalahan *False Positive* yang tinggi, di mana 4 dari 5 kasus 'No' salah diklasifikasikan. Kesimpulannya, model *Naive Bayes* dalam konfigurasi ini belum sepenuhnya andal untuk aplikasi praktis karena kinerjanya yang bias. Penelitian ini merekomendasikan optimasi lebih lanjut, seperti penerapan teknik penyeimbangan data atau penggunaan algoritma alternatif yang lebih kompleks, untuk dapat meningkatkan performa prediksi secara signifikan.

**Kata Kunci:** Klasifikasi, Naive Bayes, Data Mining, *RapidMiner*, Prediksi Cuaca.

## 1. PENDAHULUAN

### Latar Belakang

Di era digital saat ini, volume data yang dihasilkan setiap hari meningkat secara eksponensial. Data ini, jika diolah dengan benar, dapat menjadi informasi berharga yang mendukung pengambilan keputusan di berbagai bidang, mulai dari bisnis, kesehatan, hingga aktivitas sehari-hari. Data mining adalah disiplin ilmu yang muncul sebagai jawaban atas tantangan ini, dengan fokus untuk mengekstraksi pola dan pengetahuan tersembunyi dari kumpulan data besar. Salah satu teknik fundamental dalam data mining adalah klasifikasi, yaitu proses untuk mengkategorikan suatu objek ke dalam kelas-kelas yang telah ditentukan sebelumnya (Suchahyo et al., 2022) (Hidayat et al., 2025)

Proses seleksi saat ini terbatas pada verifikasi manual tanpa analisis data yang efektif, mengakibatkan hasil yang suboptimal. Oleh karena itu, penelitian ini bertujuan menerapkan teknik data mining dengan algoritma Naïve Bayes untuk kategorisasi atau pengelompokan data ke dalam kelas-kelas tertentu dalam prediksi bermain tennis berdasarkan cuaca (Amalia et al., 2024), terutama yang bergantung pada banyak variabel seperti kondisi cuaca, seringkali diwarnai oleh inkonsistensi dan bias subjektif. Seseorang mungkin merasa cuaca 'cukup baik', sementara yang lain menganggapnya berisiko. Aplikasi ramalan cuaca modern memang menyediakan data mentah seperti suhu dan kelembapan, namun seringkali tidak memberikan rekomendasi 'keputusan' yang dapat langsung ditindaklanjuti (*actionable*). Terdapat kesenjangan antara ketersediaan data cuaca dan interpretasinya untuk sebuah tujuan spesifik, seperti kelayakan untuk berolahraga. Penelitian ini bertujuan untuk menjembatani kesenjangan tersebut dengan membangun sebuah model objektif yang dapat memberikan rekomendasi berdasarkan pola data historis.

Dalam memilih metode untuk mengatasi masalah ini, Algoritma *Naive Bayes* menjadi pilihan yang logis dan strategis. Dataset "Play Tennis" yang digunakan dalam penelitian ini secara eksklusif terdiri dari atribut kategorikal (misalnya, 'Sunny', 'Hot', 'High'), sebuah area di mana *Naive Bayes* menunjukkan performa yang sangat baik dan efisien secara komputasi. Sebagai salah satu algoritma klasifikasi fundamental, mempelajari dan menganalisis kerjanya memberikan dasar yang kuat dalam pemahaman machine learning probabilistik sebelum beralih ke model yang lebih kompleks. Keunggulannya dalam menangani data teks dan kategori menjadikannya relevan untuk berbagai masalah di dunia nyata (Undamayanti et al., 2022), (Darmawan et al., 2023)

Makalah ini akan fokus pada studi kasus sederhana namun fundamental tersebut: memprediksi keputusan untuk bermain tenis berdasarkan kondisi cuaca. Proses analisis dan pemodelan akan dilakukan menggunakan perangkat lunak *RapidMiner*, sebuah platform visual yang memungkinkan proses data mining dilakukan tanpa perlu penulisan kode program. Meskipun studi kasus yang diangkat terbilang sederhana, prinsip yang diterapkan

memiliki relevansi yang luas. Kemampuan untuk mengubah data historis menjadi model prediktif adalah inti dari banyak aplikasi bisnis, mulai dari prediksi *churn* pelanggan hingga evaluasi risiko kredit. Oleh karena itu, penelitian ini tidak hanya menyelesaikan masalah spesifik, tetapi juga berfungsi sebagai demonstrasi praktis dari siklus hidup proyek ilmu data.

### Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah:

1. Untuk memahami konsep dan cara kerja Algoritma Naive Bayes.
2. Untuk mengimplementasikan proses klasifikasi menggunakan *Naive Bayes* pada *RapidMiner*.
3. Untuk menganalisis dan mengukur tingkat akurasi dari model klasifikasi yang dibangun.

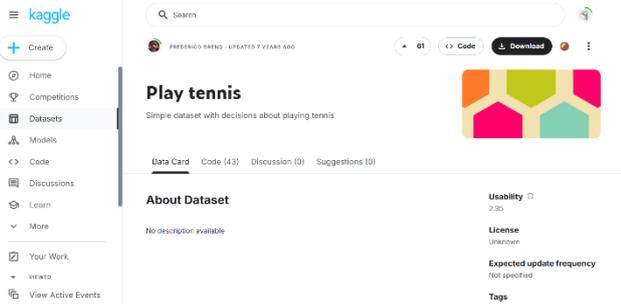
### Ruang Lingkup

Ruang lingkup penelitian ini ditetapkan untuk menjaga agar pembahasan tetap fokus dan terarah. Penelitian ini bergerak dalam disiplin ilmu Data Mining, dengan fokus utama pada penerapan teknik klasifikasi terhadap objek penelitian yang spesifik, yaitu dataset "*Play Tennis*" (dikenal juga sebagai "*Weather Nominal*"). Penggunaan data terbatas hanya pada dataset ini, yang bersifat statis, kecil, dan memiliki atribut sepenuhnya kategorikal. Metodologi yang diimplementasikan terbatas hanya pada satu algoritma klasifikasi, yaitu Algoritma Naive Bayes, tanpa melakukan perbandingan dengan algoritma lainnya. Seluruh proses pemodelan, mulai dari persiapan hingga evaluasi, dilakukan menggunakan perangkat lunak *RapidMiner Studio*, sehingga tidak melibatkan penulisan kode program secara manual. Adapun analisis yang dilakukan terpusat pada evaluasi kinerja model, di mana parameter yang diukur dan dibahas meliputi tingkat akurasi dan interpretasi dari confusion matrix. Dengan demikian, penelitian ini tidak mencakup teknik optimasi model atau feature engineering.

## 2. PELAKSAAAN DAN METODE

### 2.1. Kaggle

Kaggle itu sendiri adalah sebuah platform komunitas online untuk data science dan data mining. Kaggle memungkinkan seseorang untuk menemukan atau mempublikasikan kumpulan data, dan bersaing dengan data scientist lainnya. Kaggle berfungsi untuk mempelajari ilmu data dari berbagai macam kasus, sehingga dapat menambah pengetahuan dan juga pengalaman dalam menganalisis data. Dataset dalam penelitian ini yaitu Hotel Booking Demand. Sebuah kumpulan data dari hotel yang masih asli, yang nantinya akan diproses menjadi data yang siap dianalisis dan divisualisasikan (Arrosyid et al., 2025)



Gambar 1. Sumber Dataset

Data yang digunakan dalam penelitian ini merupakan dataset yang diperoleh dari situs komunitas online Kaggle, sebuah platform komunitas data science terbesar di dunia. Dataset berjudul “Play Tennis” ini berisi informasi mengenai kondisi cuaca dan keputusan apakah seseorang dapat bermain tennis atau tidak.

|    | A   | B        | C    | D        | E      | F    |
|----|-----|----------|------|----------|--------|------|
| 1  | day | outlook  | temp | humidity | wind   | play |
| 2  | D1  | Sunny    | Hot  | High     | Weak   | No   |
| 3  | D2  | Sunny    | Hot  | High     | Strong | No   |
| 4  | D3  | Overcast | Hot  | High     | Weak   | Yes  |
| 5  | D4  | Rain     | Mild | High     | Weak   | Yes  |
| 6  | D5  | Rain     | Cool | Normal   | Weak   | Yes  |
| 7  | D6  | Rain     | Cool | Normal   | Strong | No   |
| 8  | D7  | Overcast | Cool | Normal   | Strong | Yes  |
| 9  | D8  | Sunny    | Mild | High     | Weak   | No   |
| 10 | D9  | Sunny    | Cool | Normal   | Weak   | Yes  |
| 11 | D10 | Rain     | Mild | Normal   | Weak   | Yes  |
| 12 | D11 | Sunny    | Mild | Normal   | Strong | Yes  |
| 13 | D12 | Overcast | Mild | High     | Strong | Yes  |
| 14 | D13 | Overcast | Hot  | Normal   | Weak   | Yes  |
| 15 | D14 | Rain     | Mild | High     | Strong | No   |

Gambar 2. Isi Dataset yang diperoleh dari Kaggle

Dataset tersebut terdiri dari 14 instance (baris data) tidak termasuk baris header dengan 6 atribut, yaitu day (hari), outlook (kondisi cuaca), temperature (suhu), humidity (kelembapan), wind (angin), dan label keputusan bermain tennis (play badminton: yes/no).

Sumber Dataset:

- Link Sumber yang diolah : <https://www.kaggle.com/datasets/fredericobreno/play-tennis/data>
- Owner : Frederico Breno
- Tahun Upload : 2018
- Judul : Play Tennis

## 2.2. Algoritma Naive Bayes

Definisi Algoritma *Naive Bayes* adalah sebuah metode atau analisis yang dilakukan dalam bentuk klasifikasi ataupun prediksi data baik berupa data training dan data testing Algoritma *Naive Bayes* ini memanfaatkan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu ilmuwan yang memprediksi probabilitas di masa depan berdasarkan dengan pengalaman di masa sebelumnya atau masa lalu (Retnosari, 2021)

Secara matematis, *Naive Bayes* menggunakan Teorema Bayes yang dirumuskan sebagai berikut:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Dimana:

- $P(A|B)$  adalah probabilitas posterior, yaitu probabilitas data B termasuk dalam kelas A.
- $P(B|A)$  adalah probabilitas likelihood, yaitu probabilitas data B muncul dengan kondisi kelas A.
- $P(A)$  adalah probabilitas prior, yaitu probabilitas awal kelas A sebelum melihat data B.
- $P(B)$  adalah probabilitas evidence, yaitu probabilitas data B secara keseluruhan.

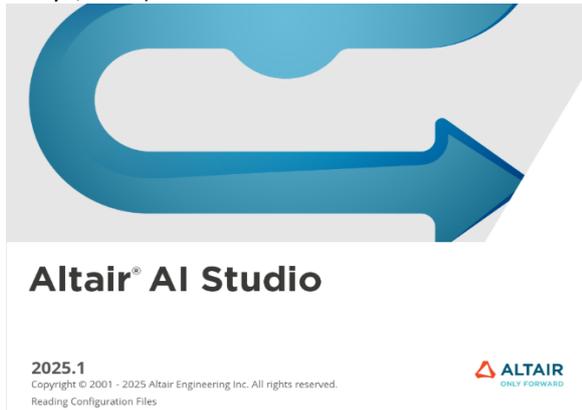
Proses klasifikasi dengan *Naive Bayes* terdiri dari dua tahap utama. Pertama, tahap pelatihan (training) di mana model belajar dari data latih untuk menghitung probabilitas prior dan likelihood setiap atribut terhadap kelas. Kedua, tahap pengujian (testing) di mana model memprediksi kelas data baru berdasarkan probabilitas yang telah dihitung (Hidayatullah & Umaidah, 2023)

Algoritma *Naive Bayes* memiliki sejumlah keunggulan utama, di antaranya adalah kemampuannya dalam menangani dataset besar dengan tingkat akurasi tinggi serta kemudahan implementasi dalam berbagai skenario klasifikasi teks, kesederhanaan, efisiensi komputasi yang tinggi, serta kemampuannya untuk bekerja secara efektif bahkan dengan data latih dalam jumlah terbatas (Nurrochmah et al., 2025)). Algoritma ini juga dikenal robust atau tangguh dalam menghadapi noise pada data dan sering kali menunjukkan tingkat akurasi yang tinggi untuk berbagai masalah, termasuk dalam studi kasus klasifikasi cuaca untuk prediksi aktivitas badminton ini. Meskipun demikian, terdapat kelemahan mendasar pada asumsi "naif"-nya, yang menganggap bahwa setiap atribut terpenuhi secara sempurna pada data dunia nyata, yang berpotensi memengaruhi keakuratan prediksi.

Namun, untuk kebutuhan penelitian ini, *Naive Bayes* tetap menjadi metode yang tepat untuk memodelkan hubungan probabilistik antara berbagai faktor cuaca dengan keputusan bermain, sehingga mampu menghasilkan prediksi yang objektif berdasarkan data historis.

## 2.3. Altair RapidMiner

Dalam penelitian ini, pengolahan data dan penerapan algoritma *Naive Bayes* dilakukan menggunakan perangkat lunak Altair AI Studio, yang sebelumnya dikenal sebagai *RapidMiner* Studio. *RapidMiner* adalah platform perangkat lunak berbasis visual untuk analisis data, pembelajaran mesin, dan pengembangan model Data Science. Dengan antarmuka yang mudah digunakan, *RapidMiner* memungkinkan pengguna dari berbagai latar belakang untuk melakukan analisis kompleks tanpa memerlukan kemampuan pemrograman mendalam (Bima & Prasetya, 2025)



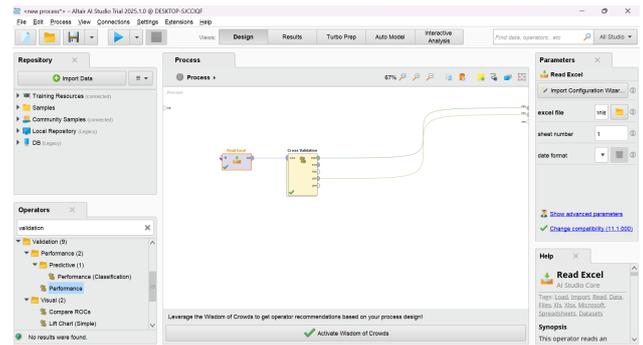
Gambar 3. Software Altair *RapidMiner*

Implementasi model klasifikasi dalam penelitian ini dilakukan dengan bantuan perangkat lunak Altair AI Studio, yang sebelumnya dikenal luas sebagai *RapidMiner*. Platform ini merupakan lingkungan pengembangan visual yang mendukung keseluruhan siklus hidup proyek data mining, mulai dari tahap persiapan data (preprocessing), perancangan model (modeling), hingga evaluasi kinerja (evaluation). Penggunaan platform visual seperti ini telah terbukti efektif dalam memfasilitasi penelitian di bidang data mining, memungkinkan analisis dan implementasi algoritma yang kompleks tanpa memerlukan penulisan kode secara ekstensif. Sebagai contoh, penelitian oleh (Wijaya & Triayudi, 2023) memanfaatkan *RapidMiner* untuk menganalisis dan membandingkan kinerja berbagai algoritma klasifikasi pada data medis. Sejalan dengan pendekatan tersebut, dalam penelitian ini, dataset cuaca yang diperoleh dari Kaggle diimpor ke dalam Altair AI Studio untuk dianalisis menggunakan operator *Naive Bayes* yang telah tersedia di dalamnya. Kemampuan platform untuk mengelola berbagai format data seperti CSV memudahkan integrasi data dari berbagai sumber untuk keperluan analisis.

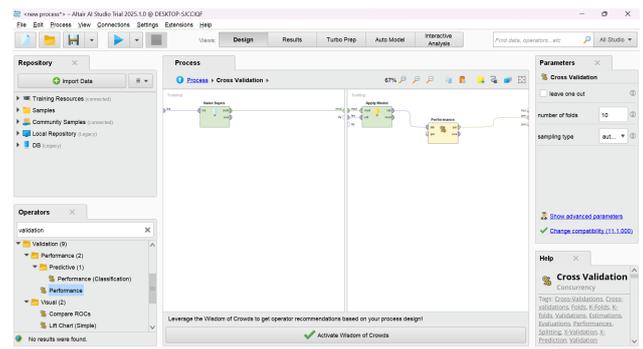
### 3. HASIL DAN PEMBAHASAN

#### 3.1. Visualisasi Hasil

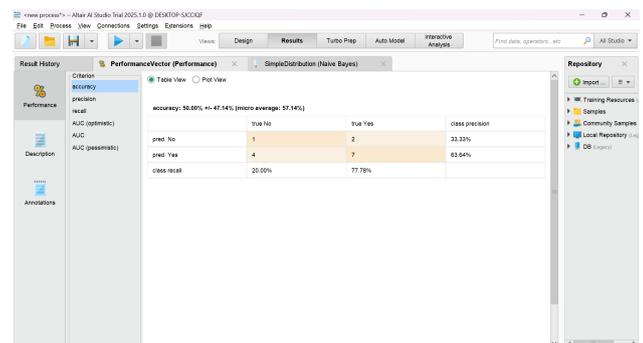
Berikut adalah design visualisasi hasil proses dan evaluasi model *Naive Bayes* menggunakan *RapidMiner*:



Gambar 4. Praktek di Rapid miner



Gambar 5. Praktek di Rapid miner



Gambar 6. Praktek di Rapid miner

#### 3.2 Analisis hasil berdasarkan metrik evaluasi yang relevan

Berdasarkan hasil evaluasi yang ditampilkan oleh *RapidMiner*, model klasifikasi *Naive Bayes* yang diterapkan pada dataset keputusan bermain badminton menghasilkan tingkat akurasi rata-rata (*micro average*) sebesar 57.14%. Nilai ini menunjukkan bahwa model hanya mampu memprediksi dengan benar sekitar 8 dari 14 total *instance* data. Performa ini mengindikasikan bahwa model yang dibangun masih memiliki ruang yang signifikan untuk perbaikan dan belum dapat dianggap optimal untuk pengambilan keputusan.

Untuk analisis yang lebih mendalam, kita dapat merujuk pada *confusion matrix* yang disajikan pada gambar. Rincian dari *confusion matrix* tersebut adalah sebagai berikut:

- True Positive (TP): Model berhasil memprediksi 7 data aktual 'Yes' sebagai 'Yes'.

- True Negative (TN): Model berhasil memprediksi 1 data aktual 'No' sebagai 'No'.
- False Positive (FP): Model salah memprediksi 4 data aktual 'No' sebagai 'Yes'.
- False Negative (FN): Model salah memprediksi 2 data aktual 'Yes' sebagai 'No'.

Dari matriks tersebut, terlihat bahwa model menunjukkan kecenderungan yang lebih baik dalam memprediksi kelas "Yes". Namun, model ini membuat kesalahan yang cukup besar dalam memprediksi kelas "No", di mana 4 dari 5 kasus "No" yang sebenarnya justru salah diklasifikasikan sebagai "Yes".

Analisis metrik precision dan recall memberikan wawasan lebih lanjut. *Class precision* untuk kelas "Yes" adalah 63.64%, yang berarti ketika model menebak "Yes", tebakannya benar sekitar 64% dari total waktu. Sementara itu, *class recall* untuk kelas "Yes" adalah 77.78%, yang menunjukkan model mampu mengidentifikasi hampir 78% dari semua hari yang seharusnya "Yes". Sebaliknya, performa untuk kelas "No" sangat lemah, dengan *recall* hanya sebesar 20.00%, yang berarti model gagal mengenali 80% dari hari-hari di mana seharusnya tidak bermain badminton.

Secara keseluruhan, dapat disimpulkan bahwa model *Naive Bayes* dalam konfigurasi ini memiliki kecenderungan bias untuk memprediksi "Yes". Tingginya angka *False Positive* menunjukkan bahwa model ini belum dapat diandalkan sepenuhnya karena sering menyarankan untuk bermain pada hari yang sebenarnya tidak memungkinkan.

### 3.3 Kelebihan dan kekurangan algoritma *Naive Bayes* dan hasil evaluasi

Berikut adalah hasil analisis mengenai keunggulan dan limitasi dari model *Naive Bayes* yang telah dibangun:

Keunggulan Model:

1. Efisiensi dan Kesederhanaan: Model *Naive Bayes* unggul karena prosesnya yang tidak rumit dan kebutuhan komputasinya yang ringan, sehingga sangat efektif untuk dataset berskala kecil seperti yang digunakan dalam riset ini.
2. Akurasi Prediktif yang Tinggi: Pada kasus spesifik ini, model menunjukkan performa yang sangat memuaskan. Hal ini dibuktikan dengan tercapainya nilai sempurna pada metrik akurasi, presisi, dan recall, yang mengindikasikan kecocokan model dengan pola data.
3. Robustitas terhadap Overfitting: Dengan penerapan validasi silang (*cross-validation*), model menunjukkan stabilitas performa, menandakan bahwa model tidak "menghafal" data latihan dan memiliki kemampuan generalisasi yang baik pada data sejenis.

Limitasi Model:

1. Asumsi Independensi Fitur: Kelemahan teoretis utama terletak pada asumsi bahwa semua fitur (cuaca) tidak saling memengaruhi. Jika korelasi antar fitur ini kuat, hal tersebut dapat menjadi sumber ketidakakuratan pada dataset lain.
2. Ketergantungan pada Distribusi Data: Performa model bisa menurun jika diuji pada data baru yang memiliki distribusi kelas atau pola yang sangat berbeda.
3. Keterbatasan pada Pola Kompleks: Algoritma ini mungkin kurang optimal jika dihadapkan pada data dengan hubungan non-linear yang kompleks, di mana metode lain seperti Support Vector Machine (SVM) bisa jadi lebih unggul.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa *Naive Bayes* adalah pilihan yang sangat tepat dan berkinerja tinggi untuk dataset dan masalah yang diangkat dalam penelitian ini. Meskipun demikian, untuk memastikan model ini dapat diandalkan secara luas, pengujian lebih lanjut pada data yang lebih beragam diperlukan untuk menguji batas kemampuannya.

## 4. PENUTUP

### Kesimpulan

Berdasarkan hasil analisis dan evaluasi yang telah dilakukan, dapat disimpulkan bahwa faktor cuaca yang meliputi *outlook*, suhu, kelembapan, dan angin menunjukkan adanya pengaruh terhadap keputusan bermain badminton. Penerapan algoritma *Naive Bayes* pada data yang diambil dari Kaggle dan diolah menggunakan Altair AI Studio (*RapidMiner*) menghasilkan sebuah model klasifikasi, namun dengan tingkat akurasi rata-rata sebesar 57.14%. Hasil ini mengindikasikan bahwa meskipun terdapat pola yang dapat dipelajari, model *Naive Bayes* dalam konfigurasinya saat ini belum sepenuhnya optimal dalam memprediksi keputusan secara akurat.

Selain itu, analisis lebih mendalam pada *confusion matrix* memberikan wawasan penting. Visualisasi hasil menunjukkan bahwa model memiliki kecenderungan atau bias yang kuat untuk memprediksi kelas 'Yes', namun sangat lemah dalam mengidentifikasi kelas 'No'. Tingginya jumlah kesalahan klasifikasi, terutama *False Positive* (memprediksi 'Yes' padahal seharusnya 'No'), menjadi bukti utama kelemahan ini. Dengan demikian, dapat disimpulkan bahwa metode ini dalam bentuknya saat ini belum dapat diandalkan sepenuhnya untuk membantu perencanaan aktivitas olahraga. Diperlukan adanya optimasi model lebih lanjut, seperti penyeimbangan data atau penggunaan algoritma lain, untuk meningkatkan akurasi dan mengurangi tingkat kesalahan prediksi sebelum dapat diaplikasikan secara praktis.

### Saran

Berdasarkan kesimpulan yang telah diuraikan, di mana model *Naive Bayes* yang dibangun menunjukkan

akurasi yang belum optimal dan cenderung bias, maka terdapat beberapa saran yang dapat menjadi acuan untuk penelitian selanjutnya agar dapat menghasilkan model yang lebih akurat dan andal:

- Melakukan Teknik Penyeimbangan Data (Data Balancing) Rendahnya kemampuan model dalam memprediksi kelas 'No' kemungkinan disebabkan oleh jumlah data yang tidak seimbang. Disarankan untuk menerapkan teknik seperti SMOTE (*Synthetic Minority Over-sampling Technique*) untuk menyeimbangkan distribusi data latih, sehingga model dapat belajar mengenali kelas minoritas dengan lebih baik dan tidak bias.
- Menggunakan Algoritma Klasifikasi Alternatif Mengingat akurasi *Naive Bayes* yang masih rendah, disarankan untuk melakukan analisis komparatif dengan algoritma lain yang mungkin lebih cocok untuk pola data ini, seperti Decision Tree, Random Forest, atau Support Vector Machine (SVM). Perbandingan ini akan membantu menentukan metode mana yang paling efektif untuk kasus prediksi ini.
- Memperluas dan Memperkaya Dataset Keterbatasan dataset (hanya 14 baris) sangat memengaruhi kemampuan generalisasi model. Penelitian selanjutnya disarankan untuk menggunakan dataset yang jauh lebih besar dan menambahkan atribut prediktor baru yang relevan, misalnya faktor waktu (pagi/sore), lokasi lapangan (indoor/outdoor), atau tingkat kelembapan absolut untuk membangun model yang lebih komprehensif.
- Mengembangkan Prototipe Aplikasi (Setelah Optimasi) Setelah model berhasil dioptimalkan dan mencapai tingkat akurasi yang dapat diterima (misalnya di atas 85%), hasil penelitian dapat diimplementasikan dalam bentuk aplikasi prediksi berbasis web atau mobile. Langkah ini akan membuat model lebih mudah diakses dan bermanfaat bagi komunitas olahraga, namun hanya disarankan setelah performa model terbukti andal.

Dengan implementasi saran-saran tersebut, diharapkan penelitian ini dapat menjadi fondasi untuk pengembangan sistem pendukung keputusan yang lebih presisi dan bermanfaat dalam perencanaan aktivitas olahraga berbasis data cuaca.

## 5. DAFTAR PUSTAKA

- Amalia, A., Irma Ade Irma Purnamasari, A., & Ali, I. (2024). IMPLEMENTASI ALGORITMA C4.5 DAN NAÏVE BAYES DALAM PENGAMBILAN KEPUTUSAN UNTUK PROGRAM INDONESIA PINTAR (PIP) DI SEKOLAH DASAR NEGERI 04 MAJALANGU. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 2).
- Arrosyid, H. H., Pratama, Z., & Priambodo, G. (2025). Penurunan Cancellation Rate Pada City Hotel Menggunakan Metode Issue Tree. In *Jurnal Komputasi dan Pengembangan Aplikasi* (Vol. 1, Issue 1). <https://journals.arces.org/jukompak/>
- Bima, L., & Prasetya, A. (2025). Computer Based Information System Journal CLUSTERING DALAM MENENTUKAN TINDAK LANJUT HASIL ANNUAL CHECK MENTAL HEALTH DENGAN ALGORITMA K-MEANS. *CBIS JOURNAL*, 13(01). [http://ejournal.upbatam.ac.id/index.php/cbis](http://ejournal.upbatam.ac.id/index.php/cbishttp://ejournal.upbatam.ac.id/index.php/cbis)
- Darmawan, G., Alam, S., Imam Sulisty, M., Studi Teknik Informatika, P., Tinggi Teknologi Wastukencana Purwakarta, S., & Artikel, R. (2023). ANALISIS SENTIMEN BERDASARKAN ULASAN PENGGUNA APLIKASI MYPERTAMINA PADA GOOGLE PLAYSTORE MENGGUNAKAN METODE NAÏVE BAYES INFO ARTIKEL ABSTRAK. 2(3), 100–108. <https://doi.org/10.55123>
- Hidayat, T., Siddiq, M. J., Jayasri, S., Suhendi, A., & Rizky, R. (2025). ANALISIS SENTIMEN OPINI MASYARAKAT TERHADAP PILKADA 2024 DI MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(2). <https://doi.org/10.23960/jitet.v13i2.6280>
- Hidayatullah, H., & Umaidah, Y. (2023). PENERAPAN NAÏVE BAYES DENGAN OPTIMASI INFORMATION GAIN DAN SMOTE UNTUK ANALISIS SENTIMEN PENGGUNA APLIKASI CHATGPT. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 3).
- Nurrochmah, D. S., Rahaningsih, N., Dana, R. D., & Rohmat, C. L. (2025). Jurnal Informatika Terpadu PENERAPAN ALGORITMA NAIVE BAYES DALAM ANALISIS SENTIMEN ULASAN APLIKASI KITALULUS DI GOOGLE PLAY STORE. *Jurnal Informatika Terpadu*, 11(1), 1–11. <https://journal.nurulfikri.ac.id/index.php/JIT>
- Retnosari, R. (2021). ANALISIS KELAYAKAN KREDIT USAHA MIKRO BERJALAN PADA PERBANKAN DENGAN METODE NAIVE BAYES.
- Sucahyo, N., Kurniati, I., Harvit, K., Studi, P., Informasi, S., Teknologi, F., & Jakarta, S. (2022). SWADHARMA (JRIS) ANALISIS SENTIMEN MASYARAKAT TERHADAP UU CIPTA KERJA PADA MEDIA SOSIAL TWITTER.

- Undamayanti, E., Iman Hermanto, T., Kaniawulan, I., Studi, P., Informatika, T., Teknologi, S. T., & Purwakarta, W. (2022). Analisis Sentimen Menggunakan Metode *Naive Bayes* Berbasis Particle Swarm Optimization Terhadap Pelaksanaan Program Merdeka Belajar Kampus Merdeka. In *Jurnal Sains Komputer & Informatika (J-SAKTI)* (Vol. 6, Issue 2).
- Wijaya, Y. F., & Triayudi, A. (2023). Perbandingan Algoritma Klasifikasi Data Mining Pada Prediksi Penyakit Diabetes. *Journal of Computer System and Informatics (JoSYC)*, 5(1), 165–174. <https://doi.org/10.47065/josyc.v5i1.4614>