



## *Comparison of Naive Bayes Classifier and Support Vector Machine Algorithms for Classifying Student Mental Health Data*

### **Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Data Kesehatan Mental Mahasiswa**

**Habib Dwi Putra<sup>1\*</sup>, Luthfia Khairani<sup>2</sup>, Delvi Hastari<sup>3</sup>**

<sup>1,2,3</sup>Program Studi Sistem Informasi, Fakultas Sains dan Teknologi,  
Universitas Islam Negeri Sultan Syarif Kasim Riau

E-mail: <sup>1</sup>12050317208@students.uin-suska.ac.id, <sup>2</sup>12050323175@students.uin-suska.ac.id,  
<sup>3</sup>12050320385@students.uin-suska.ac.id

*Corresponding Author: Habib Dwi Putra*

#### **Abstract**

*Mental well-being is essential for maximizing one's potential, understanding normal thought processes, being productive, and contributing to one's community. Currently, the mental health system is the most important aspect of education, especially in the short term. This is based on the fact that mental health contributes to the success and success of the academic community. Mental stress can negatively impact a person's ability to speak and listen to others. By using a data mining classification algorithm, mental health data is used as an input model that can be used to develop models and utilized to classify test data. The classification methods applied are the Support Vector Machines and Naïve Bayes Classifier methods. The Naive Bayes Classifier outperforms the Support Vector Machine in terms of mental bias classification with a factor of 94.37%, while the Naive Bayes Classifier outperforms the Support Vector Machine with a factor of 86.87%.*

*Keywords: Mental Health, Naive Bayes Classifier, Support Vector Machine.*

#### **Abstrak**

Kesejahteraan mental sangat penting untuk memaksimalkan potensi seseorang, memahami proses berpikir normal, menjadi produktif, dan berkontribusi pada komunitas seseorang. Saat ini, sistem kesehatan mental merupakan aspek pendidikan yang paling penting, terutama dalam jangka pendek. Hal ini didasarkan pada fakta bahwa kesehatan mental berkontribusi pada keberhasilan dan kesuksesan civitas akademika. Stres mental dapat berdampak negatif pada kemampuan seseorang untuk berbicara dan mendengarkan orang lain. Dengan menggunakan algoritma klasifikasi data mining, data kesehatan mental digunakan sebagai input model yang dapat digunakan untuk mengembangkan model dan dimanfaatkan untuk mengklasifikasikan data pengujian. Metode klasifikasi yang diterapkan adalah metode Support Vector Machines dan Naïve Bayes Classifier. Naive Bayes Classifier mengungguli Support Vector Machine dalam hal klasifikasi bias mental dengan faktor 94,37%, sedangkan Naive Bayes Classifier mengungguli Support Vector Machine dengan faktor 86,87%.

Kata Kunci: Kesehatan Mental, Naive Bayes Classifier, Support Vector Machine,

#### **1. PENDAHULUAN**

Kesehatan mental adalah komponen berbeda yang berasal dari definisinya. Sistem perawatan kesehatan mental yang baik memungkinkan orang untuk memaksimalkan potensi mereka dan mencapai tingkat kesehatan yang normal. Terlepas dari kenyataan bahwa itu adalah gangguan mental, anda tidak akan dapat mengidentifikasinya karena sebagian besar kasus pada saat itu cenderung cacat. Ada sekitar 450 orang dengan gangguan jiwa. Menurut WHO wilayah Asia Pasifik (WHO SEARO), sejumlah besar orang

menderita depresi, terutama di India (56.675.969 orang atau 4,5% dari populasi) dan Maladewa (12.739 orang atau 3,7% dari populasi). Di Indonesia terdapat 9.162.886 jiwa atau 3,7% dari jumlah penduduk [1].

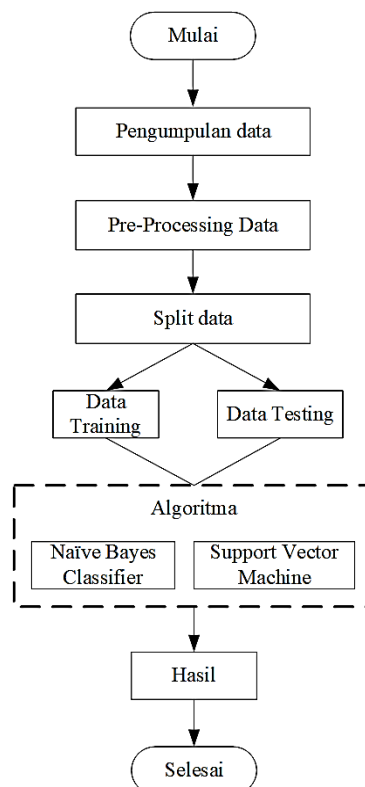
Saat ini, sistem kesehatan mental merupakan aspek pendidikan yang paling penting, terutama dalam jangka pendek. Hal ini didasarkan pada fakta bahwa kesehatan mental berkontribusi pada keberhasilan dan kesuksesan civitas akademika. Stres mental dapat berdampak negatif pada kemampuan seseorang untuk berbicara dan mendengarkan orang lain. Dengan menggunakan algoritma klasifikasi data mining, data kesehatan mental digunakan sebagai input model yang dapat digunakan untuk mengembangkan model dan dimanfaatkan untuk mengklasifikasikan data pengujian [2].

Support Vector Machines dan Naive Bayes Classifier merupakan contoh metode yang dapat digunakan untuk klasifikasi. Salah satu contoh penelitian yang telah dilakukan dengan menggunakan metode Support Vector Machines dan Naive Bayes Classifier adalah penelitian yang dilakukan oleh Shoparak, dkk (2014) untuk mendeteksi mikroaneurisma secara otomatis menggunakan pupil yang tidak melebar menggunakan metode Support Vector Machine dan Naive Bayes Classifier. Untuk mendeteksi mikroaneurisma gaib, metode Support Vector Machines menggunakan presisi yang lebih unggul dari Naive Bayes Classifier. Mayy M. Al-Tahrawi (2015) melakukan penelitian tentang kategorisasi teks bahasa Arab menggunakan regresi Support Vector Machine, Naive Bayes, dan Sistem Informasi Geografis. Hasilnya, Support Vector Machines mengungguli algoritma statistik lainnya, seperti GIS, Naive Bayes, dan regresi logistik, dalam hal akurasi [3].

Selain itu, penelitian ini menunjukkan bahwa mayoritas orang menderita gejala depresi dan kecemasan. Selain itu, ia memiliki keterampilan unik yang berkaitan dengan penelitian psikologis, dan pengetahuannya tentang subjek tersebut memberikan wawasan tentang penyakit mental [4]. Saat studi akan dilakukan, algoritma Naive Bayes Classifier dan Support Vector Machine akan diterapkan pada dataset *Student Mental Health Status* dan analisis kinerja akan dilakukan oleh algoritma kedua.

## 2. METODOLOGI PENELITIAN

Studi ini menggunakan dataset turunan Kaggle yang dikenal sebagai dataset *Student Mental*. Data di atas mengalami preprocessing, menghasilkan 101 mental data. Proses split data yang meliputi data training dan data testing dengan perbandingan 70:30 akan digunakan untuk melakukan hal tersebut. Awalnya, proses klasifikasi dilakukan menggunakan Aplikasi RapidMiner dan dua algoritma berbeda, Naive Bayes Classifier dan Support Vector Machine. Tahapan terakhir adalah menentukan suatu hasil dari klasifikasi yang dilakukan.



**Gambar 1.** Metodologi Penelitian

## 2.1 Data Mining

Data mining adalah proses pengumpulan data dari sejumlah besar data guna menemukan informasi baru yang dapat digunakan untuk meningkatkan layanan pelanggan [5].

## 2.2 Klasifikasi

Istilah "klasifikasi" mengacu pada metode yang dikenal sebagai "pembelajaran yang diawasi," yang membutuhkan penentuan jarak antara input dan atribut target. Klasifikasi yang disebutkan di atas termasuk hash yang berasal dari data [6].

## 2.3 Naive Bayes Classifier

Salah satu algoritma klasifikasi yaitu Naive Bayes Classifier (NBC) yang mana algoritma ini menggunakan probabilitas dan statistik untuk menghasilkan hipotesis, metode ini diperkenalkan oleh ilmuwan Inggris Thomas Bayes, yaitu prediksi kemungkinan masa depan berdasarkan pengalaman masa lalu [7]. Pada tahap pembelajaran, algoritma klasifikasi akan menganalisis data latih untuk membangun model klasifikasi [8]. Saat membandingkan kumpulan data tertentu dengan kumpulan data yang jauh lebih besar, Naive Bayes menemukan bahwa ada perbedaan yang signifikan antara keduanya [9]. Berikut ini adalah rumus dari NBC ditunjukkan pada persamaan 1.

$$P(X) = P(Y)P \Pi (Y) \quad (1)$$

Keterangan:

- $P(Y)$  : probabilitas data dengan vektor X pada kelas Y  
 $P(Y)$  : probabilitas awal kelas Y dan  $P(X_i | Y)$  adalah probabilitas independen kelas Y pada semua fitur dalam vektor X

## 2.4 Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah metode yang menggunakan parameter untuk mengkategorikan data berdasarkan sekumpulan parameter. SVM digunakan untuk membandingkan model dengan Scatter Plots yang membandingkan satu graf dengan persistensi data [10].

## 3. HASIL DAN PEMBAHASAN

### 3.1 Pengumpulan dan Pre-processing Data

Penelitian ini menggunakan dataset Kesehatan mental mahasiswa yang bersumber dari website Kaggle. Data yang digunakan memiliki 10 atribut yaitu *person*, *gender*, *age*, *course*, *year of study*, *Cumulative Grade Point Average (CGPA)*, *Depression*, *anxiety*, *panic attack* dan *specialist for a treatment*. Tabel 1 menunjukkan data kesehatan mental mahasiswa yang akan dilakukan pada penelitian ini.

**Tabel 1.** Data Kesehatan Mental Mahasiswa

| Person     | Gender | Age | Course             | Year of Study | CGPA        | Depression | Anxiety | Panic attack | Specialist for a Treatment |
|------------|--------|-----|--------------------|---------------|-------------|------------|---------|--------------|----------------------------|
| Person 1   | Male   | 18  | Biomedical science | year 1        | 0 - 1.99    | No         | No      | No           | No                         |
| Person 2   | Female | 19  | Human Resources    | Year 2        | 2.50 - 2.99 | No         | No      | No           | No                         |
| Person 3   | Female | 24  | BIT                | Year 3        | 3.50 - 4.00 | Yes        | Yes     | Yes          | Yes                        |
| Person 4   | Male   | 23  | Mathematics        | year 4        | 3.00 - 3.49 | No         | No      | No           | No                         |
| ...        | ...    | ... | ...                | ...           | ...         | ...        | ...     | ...          | ...                        |
| Person 101 | Female | 18  | Engineering        | year 1        | 2.00 - 2.49 | No         | No      | No           | No                         |

Sebelum diproses menggunakan algoritma, maka perlu dilakukan pre-processing data seperti transformasi dan normalisasi data. Hasil transformasi data dapat dilihat pada Tabel 2.

**Tabel 2.** Data Hasil Transformasi

| Person | Gender | Age | Course | Year of Study | CGPA | Depression | Anxiety | Panic attack | Specialist for a Treatment |
|--------|--------|-----|--------|---------------|------|------------|---------|--------------|----------------------------|
| 1      | 1      | 18  | 25     | 1             | 1    | 0          | 0       | 0            | No                         |
| 2      | 2      | 19  | 46     | 2             | 3    | 0          | 0       | 0            | No                         |

| Person | Gender | Age | Course | Year of Study | CGPA | Depression | Anxiety | Panic attack | Specialist for a Treatment |
|--------|--------|-----|--------|---------------|------|------------|---------|--------------|----------------------------|
| 3      | 2      | 24  | 3      | 3             | 5    | 1          | 1       | 1            | Yes                        |
| 4      | 1      | 23  | 52     | 4             | 4    | 0          | 0       | 0            | No                         |
| ...    | ...    | ... | ...    | ...           | ...  | ...        | ...     | ...          | ...                        |
| 101    | 2      | 18  | 1      | 1             | 2    | 0          | 0       | 0            | No                         |

Selanjutnya dilakukan normalisasi data dengan menggunakan Metode *Min-Max Normalization*. Hasil dari normalisasi data dapat dilihat pada Tabel 3 berikut ini.

**Tabel 3.** Data Hasil Normalisasi

| Id  | Gender | Age   | Course | Year of Study | CGPA  | Depression | Anxiety | Panic attack | Specialist for a Treatment |
|-----|--------|-------|--------|---------------|-------|------------|---------|--------------|----------------------------|
| 1   | 0,000  | 0,000 | 0,143  | 0,000         | 0,000 | 1,000      | 1,000   | 1,000        | No                         |
| 2   | 1,000  | 0,167 | 0,457  | 0,333         | 0,500 | 1,000      | 1,000   | 1,000        | No                         |
| 3   | 1,000  | 1,000 | 0,200  | 0,667         | 1,000 | 0,000      | 0,000   | 0,000        | Yes                        |
| 4   | 0,000  | 0,833 | 0,800  | 1,000         | 0,750 | 1,000      | 1,000   | 1,000        | No                         |
| ... | ...    | ...   | ...    | ...           | ...   | ...        | ...     | ...          | ...                        |
| 101 | 1,000  | 0,000 | 0,371  | 0,000         | 0,250 | 1,000      | 1,000   | 1,000        | No                         |

### 3.2 Klasifikasi Menggunakan NBC

Sebagai hasil dari data yang masih dalam proses normalisasi, akan dianalisis 30 data yang diaplikasikan menggunakan Rapidminer, hasil klasifikasi kesehatan mental akan ditampilkan pada Tabel 4.

**Tabel 4.** Data Hasil Uji

| Id  | Specialist for a Treatment | Prediction (Specialist for a Treatment) | Confidence (No) | Gender | Age   | ... | Confidence (Yes) |
|-----|----------------------------|---|-----------------|--------|-------|-----|------------------|
| 1   | No                         | No                                      | 0,962           | 1      | 0,667 | ... | 0,038            |
| 2   | No                         | No                                      | 1               | 0      | 0,167 | ... | 0                |
| 3   | No                         | Yes                                     | 0,002           | 0      | 0,167 | ... | 0,998            |
| 4   | No                         | No                                      | 0,903           | 1      | 0     | ... | 0,097            |
| ... | ...                        | ...                                     | ...             | ...    | ...   | ... | ...              |
| 30  | No                         | No                                      | 1               | 1      | 1     | ... | 0                |

Tahap berikutnya, memperoleh nilai *precision*, *recall* dan *accuracy*. Dengan menggunakan *tool* Rapidminer terdapat operator untuk menunjukkan hasil kinerja model, di mana ditunjukkan pada tabel 5. yaitu Confusion Matrix.

**Tabel 5.** Confusion Matrix

|              | True No | True Yes | Class Precision |
|--------------|---------|----------|-----------------|
| Pred. No     | 25      | 1        | 96,15%          |
| Pred. Yes    | 3       | 1        | 25,00%          |
| Class Recall | 89,29%  | 50,00%   |                 |
| Accuracy     |         | 86,67%   |                 |

Terdapat beberapa informasi yang dapat ditemukan pada Tabel 5, seperti data *real-time* dan data yang menunjukkan kemungkinan prediksi untuk masing-masing dari lima level dan penarikan kembali untuk masing-masing dari lima level. *Confusion Matrix* dapat digunakan untuk menguji berbagai jenis data. Untuk mengambil keputusan, metode *precision* memperhitungkan sejumlah jenis data yang berbeda dari berbagai sumber. Keakuratan model yang sedang diuji didasarkan pada ingatan yang digunakan pengklasifikasi untuk mengumpulkan data dari kumpulan data.

### 3.3 Klasifikasi Menggunakan SVM

Klasifikasi SVM dilakukan menggunakan *tool* RapidMiner, ditunjukkan pada tabel 6.

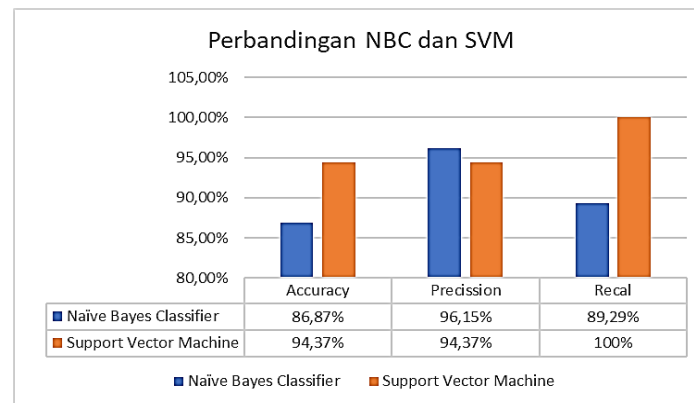
Pada Tabel 6 Dari hasil analisis diatas, menunjukkan performa SVM menghasilkan tingkat kesehatan mental siswa terjadi pada tingkat tinggi (*high level*) dengan presisi 94,37% dan *recall* sebesar 100,00% dan akurasi dari analisis model ini sebesar 94,37%.

**Tabel 6.** Confusion Matrix

|              | True No | True Yes | Class Precision |
|--------------|---------|----------|-----------------|
| Pred. No     | 67      | 4        | 94,37%          |
| Pred. Yes    | 0       | 0        | 0,00%           |
| Class Recall | 100,00% | 0,00%    |                 |
| Accuracy     |         | 94,37%   |                 |

### 3.4 Perbandingan Algoritma NBC dan SVM

Setelah data terkumpul menggunakan Algoritma NBC dan SVM, langkah selanjutnya adalah menerapkan algoritma kedua untuk mendapatkan hasil yang lebih akurat.

**Gambar 1.** Perbandingan Algoritma NBC dan SVM

## 4. KESIMPULAN

Hasil dari perbandingan Naive Bayes Classifier dan Support Vector Machine ialah untuk klasifikasi kesehatan mental mahasiswa yang didapatkan akurasi pada algoritma support vector machine dengan nilainya sebesar 94,37%, sedangkan pada algoritma Naive Bayes Classifier diperoleh akurasi sebesar 86,87%. Akurasi nilai yang tinggi menunjukkan bahwa kinerja algoritma support vector machine lebih optimal dibandingkan algoritma naive bayes classifier dalam mengklasifikasikan kesehatan mental mahasiswa, yaitu “Yes” maka mahasiswa memerlukan terapi khusus dan “No” maka mahasiswa tidak memerlukan terapi khusus.

## REFERENSI

- [1] D. Ayuningtyas, M. Misnaniarti, and M. Rayhani, “Analisis Situasi Kesehatan Mental Pada Masyarakat Di Indonesia Dan Strategi Penanggulangannya,” *J. Ilmu Kesehat. Masy.*, vol. 9, no. 1, pp. 1–10, 2018, doi: 10.26553/jikm.2018.9.1.1-10.
- [2] E. N. Wahyuni and K. Bariyyah, “Apakah spiritualitas berkontribusi terhadap kesehatan mental mahasiswa?,” *J. Educ. J. Pendidik. Indones.*, vol. 5, no. 1, p. 46, 2019, doi: 10.29210/120192334.
- [3] O. Arifin and T. B. Sasongko, “Analisa perbandingan tingkat performansi metode support vector machine dan naïve bayes classifier,” *Semin. Nas. Teknol. Inf. dan Multimed. 2018*, vol. 6, no. 1, pp. 67–72, 2018.
- [4] R. Aziz, R. Mangestuti, Y. Sholichatun, I. T. Rahayu, E. K. Purwaningtyas, and E. N. Wahyuni, “Model Pengukuran Kesehatan Mental pada Mahasiswa di Perguruan Tinggi Islam,” *J. Islam. Contemp. Psychol.*, vol. 1, no. 2, pp. 83–94, 2022, doi: 10.25299/jicop.v1i2.8251.
- [5] F. Akbar, H. Wira Saputra, A. Karel Maulaya, M. Fikri Hidayat, and Rahmaddeni, “Implementasi Algoritma Decision Tree C4.5 dan Support Vector Regression untuk Prediksi Penyakit Stroke,” vol. 2, no. October, pp. 61–67, 2022.
- [6] S. Hendrian, “Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan,” *Fakt. Exacta*, vol. 11, no. 3, pp. 266–274, 2018, doi: 10.30998/faktorexacta.v1i3.2777.
- [7] B. Delvika, S. Nurhidayarnis, and P. D. Rinada, “Comparison of Classification Between Naive Bayes and K-Nearest Neighbor on Diabetes Risk in Pregnant Women Perbandingan Klasifikasi Antara Naive Bayes dan K-Nearest Neighbor Terhadap Resiko Diabetes Pada Ibu Hamil,” vol. 2, no. October, pp. 68–75, 2022.
- [8] I. Riadi, R. Umar, and F. D. Aini, “Analisis Perbandingan Detection Traffic Anomaly Dengan Metode Naive Bayes Dan Support Vector Machine (Svm),” *Ilk. J. Ilm.*, vol. 11, no. 1, pp. 17–24,

- 2019, doi: 10.33096/ilkom.v11i1.361.17-24.
- [9] Nurhidayati, Yahya, Fathurrahman, L. . Samsu, and W. Amnia, “Implementasi Algoritma Naive Bayes Untuk Klasifikasi Penerima Beasiswa ( Studi Kasus Universitas Hamzanwadi ) Setiap warga negara Indonesia berhak mendapatkan pendidikan sebagaimana terrtuang dalam UUD 1945 pasal 31 ayat 1-2 , namun pada kenyataannya tida,” *Infotek J. Inform. dan Teknol.*, vol. 6, no. 1, pp. 177–188, 2023.
- [10] A. Z. Macfud, A. P. Kusuma, and W. D. Puspitasari, “ANALISIS ALGORITMA NAIVE BAYES CLASSIFIER ( NBC ),” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 87–94, 2023.