



# Peningkatan Performa Naive Bayes dengan Fitur Chi-Square pada Analisis Sentimen Komentar Pengguna Aplikasi Netflix

Pareza Alam Jusia<sup>1</sup>, Riza Pahlevi<sup>1</sup>, Daniel Sintong Pardamean Simanjuntak<sup>2</sup>, Jasmir<sup>3,\*</sup>

<sup>1</sup> Fakultas Ilmu Komputer, Program Studi Teknik Informatika Universitas Dinamika Bangsa, Kota Jambi, Indonesia

<sup>2</sup> Fakultas Ilmu Komputer, Program Studi Sistem Informasi Universitas Dinamika Bangsa, Kota Jambi, Indonesia

<sup>3</sup> Fakultas Ilmu Komputer, Program Studi Sistem Komputer Universitas Dinamika Bangsa, Kota Jambi, Indonesia

Email: <sup>1</sup>parezaalam@gmail.com, <sup>2</sup>rizapahlevikuliah@gmail.com, <sup>3</sup>daniel.sintong.87@gmail.com, <sup>4,\*</sup>ijay\_jasmir@yahoo.com  
Email Penulis Korespondensi: ijay\_jasmir@yahoo.com

**Abstrak**—Penelitian ini membahas tentang analisis sentimen menggunakan algoritma Naive Bayes dengan Chi-Square. Tujuan dari penelitian ini adalah untuk mengetahui pengaruh pemilihan fitur Chi-Square terhadap kinerja algoritma Naive Bayes dalam menganalisis sentimen dokumen. Data penelitian diambil dari komentar pengguna Aplikasi Netflix. Pengujian dilakukan dengan menganalisis sentimen dokumen dengan dan tanpa pemilihan fitur Chi-Square. Selanjutnya dievaluasi menggunakan metode akurasi, presisi, dan recall. Hasil dari penelitian ini adalah penambahan fitur CS pada NB secara signifikan memperbaiki semua metrik evaluasi, terutama recall dan F1-score, yang menunjukkan bahwa fitur tambahan membantu meningkatkan kemampuan model dalam memahami data. Sebelum menggunakan Chi-Square hasil akurasi adalah 80,66% dan setelah menggunakan Chi-Square akurasi meningkat menjadi 86,22%. Terjadi peningkatan sebesar 5,56%. Kombinasi NB + CS dengan split 70:30 memberikan hasil terbaik, sehingga menjadi pilihan yang optimal.

**Kata Kunci:** Analisis Sentimen; Naive Bayes; Chi-Square; Fitur Ekstraksi; Peningkatan Performa.

**Abstract**—This study discusses sentiment analysis using the Naive Bayes algorithm with Chi-Square. The purpose of this study is to determine the effect of Chi-Square feature selection on the performance of the Naive Bayes algorithm in analyzing document sentiment. The research data was taken from Netflix Application user comments. Testing was carried out by analyzing document sentiment with and without Chi-Square feature selection. Furthermore, it was evaluated using the accuracy, precision, and recall methods. The results of this study are that the addition of CS features to NB significantly improves all evaluation metrics, especially recall and F1-score, indicating that additional features help improve the model's ability to understand data. The combination of NB + CS with a 70:30 split gives the best results, making it the optimal choice.

**Keywords:** Sentiment Analysis; Naive Bayes; Chi-Square; Feature Extraction; Performance Improvement

## 1. PENDAHULUAN

Analisis sentimen atau opinion mining merupakan suatu proses otomatis untuk memahami, mengekstraksi, dan mengolah data tekstual guna memperoleh informasi sentimen yang terkandung dalam suatu pernyataan opini [1]. Tujuan dari analisis sentimen ini adalah untuk mengidentifikasi apakah opini seseorang terhadap suatu isu atau topik berita cenderung negatif, positif, atau netral, sehingga kumpulan opini tersebut dapat memberikan informasi yang bermanfaat [2][3]. Informasi yang ada dalam dokumen tersebut merupakan data teks yang tidak memiliki pola yang jelas. Menganalisis sentimen komentar berbahasa Indonesia memiliki tantangan tersendiri. Bahasa Indonesia memiliki karakteristik yang unik, seperti penggunaan singkatan, slang, ejaan yang tidak standar, dan struktur kalimat yang fleksibel, yang seringkali tidak terakomodasi dengan baik oleh model analisis sentimen yang umum digunakan. Selain itu, sentimen dalam komentar seringkali bersifat implisit atau tersirat, membutuhkan pemahaman kontekstual yang mendalam.

Text mining diperlukan untuk menangani teks yang tidak memiliki pola dan memiliki karakter yang unik tersebut [4]. Salah satu kegiatan utama dalam text mining adalah klasifikasi atau kategorisasi. Text mining merupakan suatu teknik yang dapat digunakan untuk mengklasifikasikan dokumen, dimana teknik ini merupakan variasi dari data mining yang bertujuan untuk menemukan pola-pola yang menarik dari kumpulan data tekstual yang besar [5][6]. Proses klasifikasi teks telah dilakukan dengan menggunakan berbagai metode, antara lain Naive Bayes [7], K-Nearest Neighbor [8], dan Decision Trees [9].

Penelitian mengenai analisis sentimen menggunakan metode Naive Bayes sebelumnya telah dilakukan oleh Harpizon [10] yang membahas tentang Analisis Sentimen YouTube Kuliah UAS dengan Naive Bayes dengan akurasi sebesar 87%. Sementara itu, berdasarkan penelitian oleh Elly Indrayuni [11] yang membahas tentang Klasifikasi Teks Ulasan Produk Kosmetik Indonesia dengan Naive Bayes dengan akurasi sebesar 90,5%. Keunggulan metode Naive Bayes adalah kesederhanaannya, tetapi tetap mampu memberikan akurasi yang tinggi. Salah satu keunggulan utama penggunaan metode ini adalah dibutuhkannya data latih yang relatif kecil untuk mengestimasi parameter-parameter yang diperlukan, seperti rata-rata dan varians variabel, dalam proses klasifikasi [12][13].

Metode Naive Bayes memang sangat sederhana dan efisien, namun di sisi lain metode ini sangat sensitif terhadap seleksi fitur. Menurut Fitria [14], tantangan utama dalam klasifikasi teks adalah dimensi ruang fitur yang tinggi, terutama pada teks yang memiliki puluhan ribu fitur. Sebagian besar fitur tersebut tidak relevan dan tidak berguna bagi klasifikasi teks, bahkan dapat menurunkan tingkat akurasi. Oleh karena itu, pemilihan fitur yang tepat menjadi sangat penting.

Penelitian yang dilakukan oleh Oky Irnawati [15] menunjukkan bahwa seleksi fitur merupakan langkah krusial dalam klasifikasi teks dan berdampak langsung pada kinerja klasifikasi. Diketahui bahwa klasifikasi teks menghadapi permasalahan terkait dengan dimensi data yang tinggi [16][17]. Hal ini disebabkan oleh representasi dokumen sebagai



kumpulan kata (bag of words), di mana setiap kata dalam dokumen dianggap independen satu sama lain. Oleh karena itu, pemilihan fitur yang tepat menjadi salah satu cara untuk meningkatkan kinerja klasifikasi.

Salah satu fitur yang dapat meningkatkan kinerja klasifikasi adalah fitur Chi-Square. Said Bahassine[18] mengungkapkan bahwa dalam melakukan klasifikasi suatu dokumen, Chi-Square merupakan salah satu seleksi fitur supervised yang dapat menghilangkan banyak fitur tanpa mengurangi tingkat akurasi.

Pada penelitian sebelumnya, Muhammad Biki Hamzah[19] melakukan penelitian tentang analisis sentimen review film dengan naive bayes dan chi-square, memperoleh peningkatan kinerja dari 81,39% menjadi 85,37%. Berikutnya, Anisa Falasari[20] meneliti analisis sentimen review Amazon menggunakan naive bayes dan chi-square, dan memperoleh peningkatan kinerja dari 82% menjadi 83%. Selanjutnya, Nurhadi Wijaya[21] juga melakukan penelitian menggunakan naive bayes dan chi-square dan berhasil memperoleh peningkatan nilai kinerja klasifikasi sebesar 89,59%. Oleh karena itu, pada penelitian ini, penulis memilih menggunakan metode Chi-Square sebagai metode seleksi fitur. Kontribusi penelitian ini adalah meningkatkan nilai kinerja klasifikasi metode naive bayes dari teks analisis sentimen komentar pengguna aplikasi Netflix menggunakan fitur chi-square. Penelitian ini dilakukan untuk membuktikan pengaruh seleksi fitur Chi-Square terhadap kinerja algoritma Naïve Bayes dalam menganalisis sentimen suatu dokumen.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Metodologi penelitian ini kami susun dalam bentuk rangkaian kegiatan atau kerangka kerja penelitian, seperti terlihat pada gambar 1.

#### a. Studi Literatur

Pada tahap ini penulis melakukan analisis terhadap permasalahan yang berkaitan dengan ekstraksi fitur untuk meningkatkan nilai evaluasi kinerja klasifikasi. Kajian pustaka juga dilakukan dengan tujuan untuk melihat apakah model yang akan dikembangkan dapat meningkatkan nilai evaluasi kinerja klasifikasi..

#### b. Pemilihan Dataset dan Crawling.

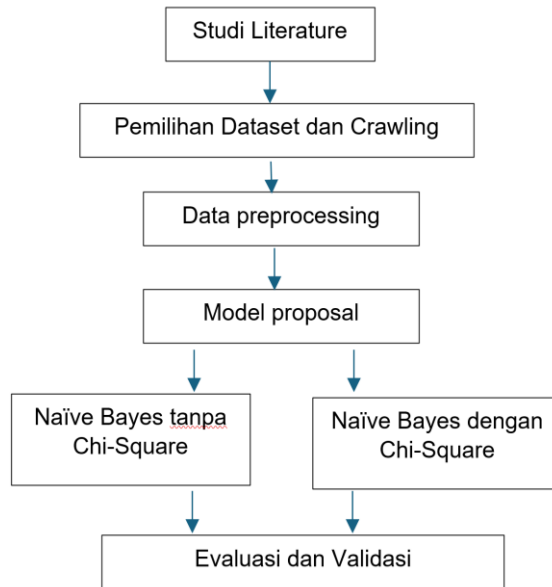
Pemilihan dataset ditentukan berdasarkan data yang akan diolah dan mencari data yang tersedia serta memperoleh data tambahan yang dibutuhkan, kemudian mengintegrasikan data tersebut ke dalam dataset. Data yang digunakan dalam penelitian ini diambil melalui komentar langsung dari pengguna Aplikasi Netflix menggunakan bahasa pemrograman Python dengan teknik Web Scrapping. Dataset diperoleh melalui proses pengumpulan data yang dilakukan dengan cara memanfaatkan library Google Play Scraper Python. Untuk melakukan scrapping, pertama-tama diperlukan ID dari aplikasi yang ingin diambil datanya. Dalam kasus ini, Netflix memiliki ID 'com.netflix.mediaclient'. Berikutnya, pemilihan bahasa pada review menjadi langkah penting, dimana penelitian ini hanya mempertimbangkan review dalam bahasa Indonesia. Setelah pemilihan bahasa, pemilihan review dilakukan berdasarkan skor. Dalam penelitian ini, review yang diambil memiliki rentang skor 1 hingga 5. Selanjutnya, urutan review yang digunakan adalah Most Relevant. Jumlah data yang akan diambil juga perlu ditentukan. Data yang diperoleh memiliki beberapa atribut, antara lain: reviewId, username, userImage, content, score, thumbsUpCount, reviewCreatedVersion, at, replyContent, answeredAt, dan appVersion. Namun, tidak semua atribut tersebut diperlukan untuk penelitian ini. Oleh karena itu, atribut yang tidak relevan atau tidak digunakan dihilangkan untuk menyederhanakan data. Ada 4 atribut yang akan digunakan, yaitu Username, Score, Date dan Content.

#### c. Data preprocessing.

Pemilihan data dilakukan. Data dibersihkan dan diubah ke dalam bentuk yang diinginkan sebelum dibuat model. Dataset yang digunakan hanya sebanyak 900 record. Dataset ini dalam tahap preprocessing melalui 3 proses yaitu: Tokenization, Stopwords Removal dan Stemming serta Labeling. Tokenisasi adalah proses memecah kalimat atau paragraf menjadi unit-unit yang lebih kecil, yang disebut token. Dalam analisis sentimen, token biasanya adalah kata atau frasa individual. Contoh: "Film ini sangat menghibur" rightarrow ["Film", "ini", "sangat", "menghibur"]. Stop words adalah kata-kata umum yang sering muncul dalam teks tetapi biasanya tidak membawa makna sentimen yang signifikan (misalnya, "yang", "dan", "di", "ke", "adalah"). Menghapus kata-kata ini dapat mengurangi dimensi data dan fokus pada kata-kata yang lebih bermakna. Contoh: "Film ini sangat bagus sekali" rightarrow ["Film", "sangat", "bagus"] (setelah "ini" dan "sekali" dihapus sebagai stop words). Stemming: Memotong imbuhan (awalan, sisipan, akhiran) dari sebuah kata untuk mendapatkan kata dasar, seringkali tanpa mempertimbangkan arti morfemnya. Contoh: "menonton", "ditonton", "tontonan" rightarrow "tonton". Kemudian tahap labeling adalah tahap menentukan kelas. Tahap labeling dilakukan secara manual.

#### d. Model Proposal.

Data yang dipelajari dan dianalisis kemudian dikelompokkan ke dalam variabel-variabel yang saling berhubungan, kemudian dibuat suatu model yang sesuai dengan jenis datanya. Pembagian data menjadi data latih dan data uji juga diperlukan untuk pembuatan model. Dalam penelitian ini kami menggunakan pembagian data 80:20 dan 70:30



Gambar 1. Kerangka Kerja Penelitian

e. Experiment

Pada bagian ini kami melakukan percobaan dengan dua cara yaitu percobaan menggunakan naive bayes tanpa fitur chi-square dan percobaan menggunakan naive bayes dengan fitur chi-square. Percobaan ini dilakukan untuk melihat hasil berupa aturan yang akan digunakan dalam pengambilan keputusan pada hasil penelitian. Proses percobaan ini menggunakan pemrograman python. Untuk pengujian model digunakan dataset komentar pengguna aplikasi Netflix. Seleksi fitur atau pemilihan kata dilakukan untuk mereduksi fitur-fitur yang tidak relevan dalam proses klasifikasi oleh NBC. Terdapat beberapa metode untuk penyeleksian fitur yaitu Mutual Information (MI), Chi-Square ( $\chi^2$ ), dan yang umum digunakan adalah frequency-based. Dalam penelitian ini digunakan seleksi fitur dengan *Chi-Square* ( $\chi^2$ ). Penyeleksian fitur *Chi-Square* dilakukan dengan cara mengurutkan setiap fitur berdasarkan hasil seleksi fitur *Chi-Square* dari nilai yang terbesar hingga nilai yang terkecil. Seleksi fitur atau pemilihan kata dilakukan dengan metode *Chi-Square*. Pengambilan keputusan dilakukan berdasarkan nilai chi ( $\chi^2$ ) dari masing-masing kata. Kata yang memiliki nilai chi ( $\chi^2$ ) lebih besar dari nilai kritis pada taraf nyata  $\alpha$  0.25, 0.1, dan 0.05 yaitu 1.323, 2.706, dan 3.841 adalah kata yang dipilih.

Penghitungan nilai *Chi-Square* pada setiap kata  $t$  yang muncul pada setiap kelas  $c$  dapat dibantu dengan menggunakan tabel kontingensi pada Tabel 1

Tabel 1. Tabel kontingensi Chi-Square

| Kata | Kelas |    |
|------|-------|----|
|      | C     | C* |
| t    | A     | B  |
| t*   | C     | D  |

Nilai yang terdapat pada Tabel 1 merupakan nilai frekuensi observasi dari suatu kata terhadap kelas dengan keterangan sebagai berikut :

A = banyaknya dokumen pada kelas c yang memuat kata t

B = banyaknya dokumen yang tidak ada di kelas c namun memuat kata t. C = banyaknya dokumen yang ada di kelas c namun tidak memiliki kata t D = banyaknya dokumen yang tidak ada di kelas c dan tidak memuat kata t

Pada tahap ini, tiap kata yang diperoleh dihitung menggunakan persamaan sebagai berikut :

$$\chi^2(t,c) = \frac{N(AD-CB)^2}{(A+C)B+(A+B)(C+D)} \tag{1}$$

Terlihat bahwa  $\chi^2(t,c)$  adalah nilai statistik Chi-Kuadrat yang dihitung dari data Anda. Notasi (t,c) kemungkinan merujuk pada "test calculated" (nilai uji yang dihitung) atau bisa juga merujuk pada konteks tertentu dari tabel (misalnya, terkait dengan baris dan kolom). N adalah Ini adalah total keseluruhan observasi atau jumlah sampel. Dihitung dengan menjumlahkan semua sel dalam tabel:  $N=A+B+C+D$ . Sementara A adalah jumlah observasi pada baris pertama, kolom pertama. B adalah jumlah observasi pada baris pertama, kolom kedua. C adalah jumlah observasi pada baris kedua, kolom pertama. Dan D adalah jumlah observasi pada baris kedua, kolom kedua.

Penyebut dalam rumus terdiri dari (A+C) yang merupakan total observasi di kolom pertama. (B+D) merupakan total observasi di kolom kedua. (A+B) merupakan total observasi di baris pertama. Dan (C+D) merupakan total observasi di baris kedua. Bagian pembilang  $N(AD-CB)^2$ : (AD-CB): Ini adalah determinan dari matriks 2x2 yang dibentuk oleh



A, B, C, dan D. Perbedaan ini mengukur sejauh mana data Anda menyimpang dari kondisi independen (tidak ada hubungan). Mengkuadratkannya memastikan nilainya selalu positif. Mengalikannya dengan N memberikan bobot berdasarkan ukuran sampel keseluruhan.

Kemudian diambil kata-kata yang memiliki nilai *chi square* lebih besar dari nilai kritis sesuai taraf nyata. Semakin kecil taraf nyata  $\alpha$  dan semakin besar nilai kritis, maka jumlah fitur yang dihasilkan akan semakin sedikit dan akan mengurangi beban database. Sehingga menghasilkan kata-kata yang merupakan ciri dari suatu kategori tertentu yang selanjutnya akan digunakan dalam proses klasifikasi

Salah satu metode klasifikasi yang dapat digunakan adalah algoritma Naïve Bayes yang sering disebut dengan Naive Bayes Classifier (NBC). Naive Bayes Classifier merupakan salah satu metode machine learning yang menggunakan perhitungan probabilitas. Algoritma Naive Bayes Classifier merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data pengujian pada kategori yang paling tepat. Kelebihan metode NBC adalah sederhana tetapi memiliki akurasi yang tinggi. (Oktasari, et al., 2016). Untuk menghitung probabilitas kondisional secara efisien, NB mengasumsikan semua fitur independen. Asumsi naif ini dapat memberikan *trade-off* yang wajar antara kinerja dan biaya komputasi.

#### f. Evaluasi dan Validasi

Dalam suatu penelitian dilakukan evaluasi terhadap model yang telah dibangun untuk mengetahui tingkat keakuratan model. Setelah model selesai dilatih, kinerjanya harus dievaluasi menggunakan data uji yang belum pernah dilihat sebelumnya. Metrik evaluasi yang umum dalam analisis sentimen meliputi:

1. Akurasi (Accuracy): Proporsi prediksi yang benar dari total prediksi.
2. Presisi (Precision): Dari semua yang diprediksi positif, berapa banyak yang benar-benar positif.
3. Recall (Sensitivitas): Dari semua yang seharusnya positif, berapa banyak yang berhasil diprediksi positif.
4. F1-Score: Rata-rata harmonik dari presisi dan *recall*, berguna ketika ada ketidakseimbangan kelas.

### 3. HASIL DAN PEMBAHASAN

Bagian ini memberikan gambaran hasil dan pertimbangan yang diperoleh dari percobaan yang dilakukan sesuai dengan kerangka penelitian yang diuraikan pada bagian sebelumnya. Percobaan berkisar pada penilaian data teks media sosial menggunakan metode naive bayes dengan fitur chi-square, dengan pembagian data 80:20. Pengujian yang dilakukan pada penelitian ini adalah pengujian Naive Bayes.

Tabel 2 merupakan hasil pengujian analisis sentimen komentar pengguna Aplikasi Netflix menggunakan algoritma Naive Bayes tanpa menggunakan fitur. Hasil pengujian disimpan dalam matriks konfusi dengan masing-masing hasil evaluasi. Dapat dilihat bahwa hasil false positive = 109 dan false negative = 98. Nilai-nilai tersebut tergolong sangat tinggi sehingga menghasilkan nilai akurasi yang rendah.

**Tabel 2.** Confusion Matrix Naïve Bayes dengan split data 80:20

|                 |             | Actual Class |            |
|-----------------|-------------|--------------|------------|
|                 |             | Class = Yes  | Class = No |
| Predicted Class | Class = Yes | TP = 373     | FP = 109   |
|                 | Class = No  | FN = 98      | TN = 320   |

Tabel 3 merupakan hasil pengujian analisis sentimen komentar pengguna Aplikasi Netflix menggunakan algoritma Naive Bayes masih tanpa menggunakan fitur, dengan komposisi data split 70:30. Hasil pengujian disimpan dalam matriks konfusi dengan masing-masing hasil evaluasi. Terlihat bahwa hasil false positive = 92 dan false negative = 82. Nilai tersebut masih tinggi, sedikit lebih baik dibandingkan dengan komposisi 80:20.

**Tabel 3.** Confusion Matrix Naïve Bayes dengan split data 70:30

|                 |             | Actual Class |            |
|-----------------|-------------|--------------|------------|
|                 |             | Class = Yes  | Class = No |
| Predicted Class | Class = Yes | TP = 445     | FP = 92    |
|                 | Class = No  | FN = 82      | TN = 281   |

Tabel 4 merupakan hasil pengujian analisis sentimen komentar pengguna Aplikasi Netflix menggunakan algoritma Naive Bayes dan menggunakan fitur Chi-Square dengan komposisi data split 80:20. Hasil pengujian disimpan dalam matriks konfusi dengan setiap hasil evaluasi. Terlihat bahwa hasil false positive = 99 dan false negative = 82. Nilai-nilai tersebut juga berada pada area ideal untuk meningkatkan nilai evaluasi kinerja klasifikasi.

**Tabel 4.** Confusion Matrix Naïve Bayes dengan Chi-Square dengan split data 80:20

|                 |             | Actual Class |            |
|-----------------|-------------|--------------|------------|
|                 |             | Class = Yes  | Class = No |
| Predicted Class | Class = Yes | TP = 509     | FP = 99    |
|                 | Class = No  | FN = 82      | TN = 221   |

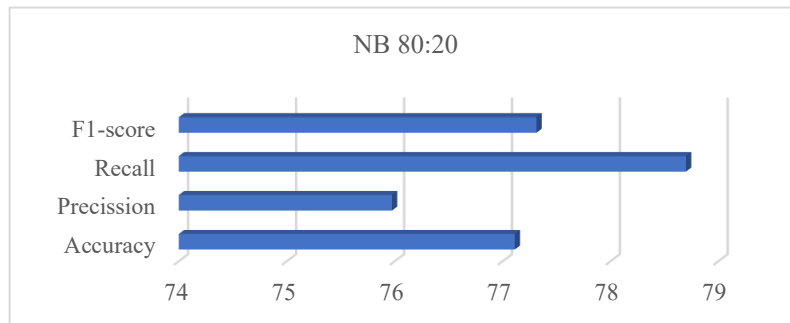


Tabel 5 merupakan hasil pengujian analisis sentimen komentar pengguna Aplikasi Netflix menggunakan algoritma Naive Bayes dan menggunakan fitur Chi-Square dengan komposisi data split 70:30. Hasil pengujian disimpan dalam matriks konfusi dengan setiap hasil evaluasi. Terlihat bahwa hasil false positive = 81 dan false negative = 71. Nilai-nilai tersebut juga berada pada area ideal untuk meningkatkan nilai evaluasi kinerja klasifikasi.

**Table 5.** Confusion Matrix Naïve Bayes dengan Chi-Square dengan split data 70:30

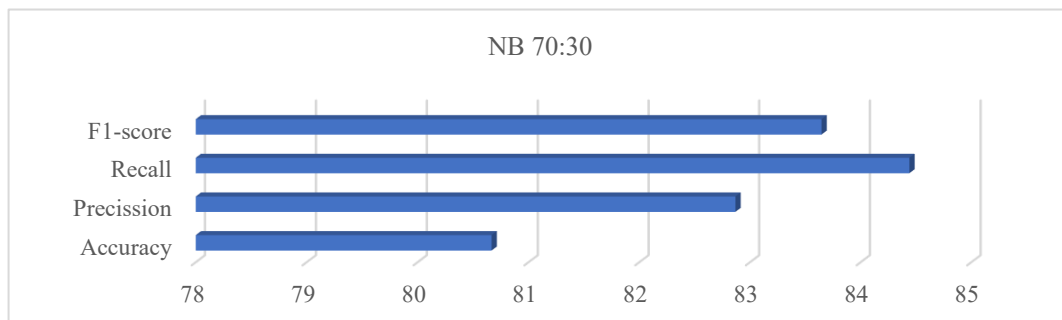
| Predicted Class |             | Actual Class |            |
|-----------------|-------------|--------------|------------|
|                 |             | Class = Yes  | Class = No |
| Class           | Class = Yes | TP = 471     | FP = 81    |
|                 | Class = No  | FN = 71      | TN = 287   |

Gambar 2 memperlihatkan perbandingan antara data aktual dengan data prediksi. Setelah dilakukan pemodelan, dapat dilihat bahwa akurasi tertinggi diperoleh dari pemodelan Naive Bayes dengan pemilihan fitur Chi Square pada skenario 80:20 dengan mendapatkan nilai akurasi tertinggi sebesar 96,97%. Ketika model digunakan untuk memprediksi seluruh dataset. Accuracy (77,11%) Menunjukkan bahwa sebanyak 77,11% dari total prediksi model adalah benar. Cukup baik, namun masih terdapat ruang untuk perbaikan, terutama jika dibandingkan dengan model lain yang mungkin lebih optimal dengan teknik feature selection. Precision (75,97%) Mengindikasikan bahwa dari seluruh prediksi positif yang dilakukan oleh model, sekitar 75,97% benar-benar positif. Nilai ini menandakan bahwa model menghasilkan cukup banyak false positives, yang bisa berdampak pada keputusan berbasis model (misalnya dalam klasifikasi penyakit, spam, dll). Recall (78,70%) Menunjukkan bahwa dari semua data yang sebenarnya positif, model mampu mendeteksi 78,70% dengan benar. Artinya, masih terdapat sekitar 21,3% data positif yang gagal terdeteksi (false negatives). F1-score (77,31%) Merupakan harmonik rata-rata dari precision dan recall. Nilai F1 yang cukup seimbang menunjukkan bahwa model memiliki keseimbangan antara kemampuan menghindari false positives dan false negatives, meskipun masih dalam kategori menengah. Model Naive Bayes dengan split data 80:20 memberikan performa cukup stabil, namun belum optimal. Recall lebih tinggi daripada precision, menandakan model lebih fokus pada menemukan semua yang positif, meskipun ada risiko salah deteksi positif. Cocok digunakan untuk aplikasi yang membutuhkan sensitivitas tinggi, tetapi perlu diwaspadai untuk konteks yang memerlukan akurasi prediksi positif yang ketat.



**Gambar 2.** Perbandingan nilai evaluasi Naïve Bayes tanpa Fitur dengan split data 80:20

Gambar 3 Model Naive Bayes dengan pembagian data 70:30 menunjukkan performa yang cukup baik dan seimbang. Nilai precision, recall, dan F1-score yang berada di atas 80% menunjukkan kinerja yang konsisten dalam klasifikasi data. Model ini cocok digunakan untuk aplikasi di mana ketepatan dan kelengkapan deteksi sama-sama penting, seperti dalam sistem diagnosis penyakit atau analisis sentimen. Jika dibandingkan dengan pembagian data 80:20 sebelumnya (yang hanya mencapai 77% akurasi), pembagian 70:30 menghasilkan model yang lebih optimal.

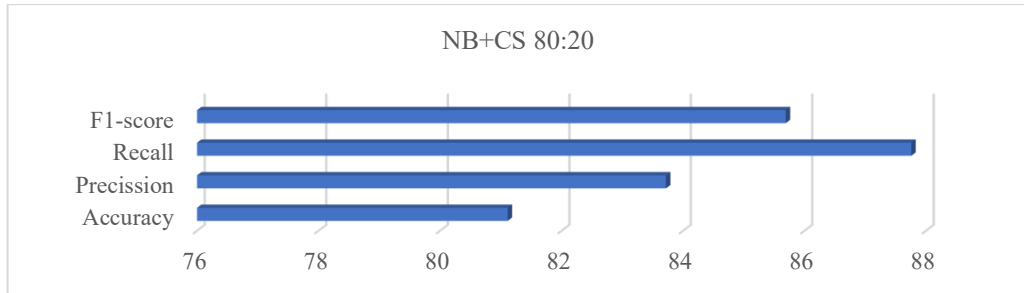


**Gambar 3.** Perbandingan nilai evaluasi Naïve Bayes tanpa Fitur dengan split data 70:30.

Gambar 4 menjelaskan Accuracy: 81,11% artinya dari total prediksi yang dibuat oleh model adalah benar (baik positif maupun negatif). Ini adalah ukuran umum performa, tetapi bisa menyesatkan jika data tidak seimbang (misalnya, satu kelas jauh lebih banyak daripada yang lain). Precision: 83,72% Precision mengukur seberapa banyak dari hasil positif yang diprediksi oleh model benar-benar positif. Nilai tinggi ini menunjukkan bahwa false positives cukup rendah; model



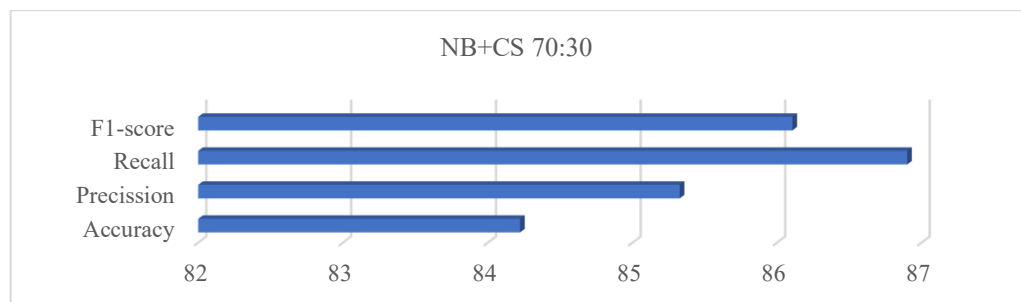
cukup andal dalam "mengatakan sesuatu adalah positif". Recall: 87,76%. Recall mengukur seberapa banyak dari seluruh kasus positif yang sebenarnya dapat ditangkap oleh model. Nilai tinggi menunjukkan bahwa model sangat baik dalam menemukan semua data positif, artinya false negatives rendah. F1-score: 85,69%. F1-score adalah harmonic mean dari precision dan recall. Skor ini menunjukkan keseimbangan yang baik antara false positives dan false negatives. Jadi Model Naive Bayes dengan CS pada skenario 80:20 menunjukkan performa baik di semua metrik. Recall lebih tinggi dari precision, artinya model lebih agresif dalam mendeteksi kelas positif (kemungkinan ada trade-off dengan meningkatnya false positives). F1-score yang tinggi menunjukkan bahwa model cukup seimbang dan dapat diandalkan untuk skenario klasifikasi ini.



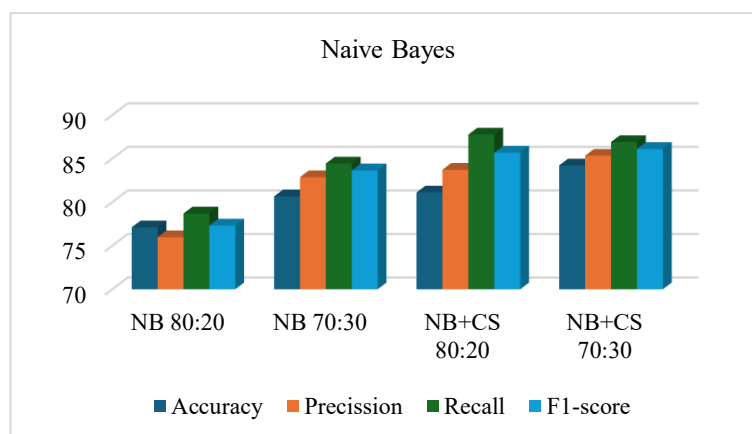
Gambar 4. Perbandingan nilai evaluasi Naïve Bayes dengan Fitur Chi-Square dengan split data 80:20

Gambar 5 menjelaskan bahwa nilai Accuracy: 84,22% Menunjukkan bahwa 84,22% dari semua prediksi model adalah benar. Kelemahan: Tidak cocok bila dataset tidak seimbang (misalnya, lebih banyak data negatif daripada positif). Precision: 85,33%. Dari seluruh prediksi positif, 85,33% adalah benar-benar positif. Kegunaan: Penting ketika false positive lebih merugikan, misalnya dalam spam filtering. Recall: 86,90%. Dari seluruh data yang sebenarnya positif, model berhasil menangkap 86,90%. Kegunaan: Penting saat kita ingin menghindari false negative, misalnya dalam deteksi penyakit. F1-score: 86,11%. Rata-rata harmonis dari precision dan recall. Jadi Nilai F1-score yang tinggi (86,11%) menunjukkan keseimbangan yang baik antara precision dan recall. Recall lebih tinggi daripada precision, artinya model lebih "berani" dalam mengklasifikasikan sebagai positif, meskipun kadang salah. Secara keseluruhan, performa model cukup baik, terutama jika dibandingkan dengan baseline atau model sederhana lainnya.

Pembagian data 70% train dan 30% test memberikan hasil evaluasi yang lebih representatif, karena jumlah data uji lebih besar daripada pada skenario 80:20, sehingga generalizability model tampak lebih baik. Jadi Model Naive Bayes dengan skenario pembagian data 70:30 menunjukkan performa yang lebih baik dan lebih seimbang dibanding skenario 80:20, terutama dalam hal accuracy dan F1-score. Hal ini dapat disebabkan oleh lebih banyak data uji, yang memberikan gambaran kinerja model yang lebih realistis.



Gambar 5. Perbandingan nilai evaluasi Naïve Bayes dengan Fitur Chi-Square dengan split data 70:30



Gambar 6. Perbandingan nilai naïve bayes dengan dan tanpa fitur chi-square.



Gambar 6 merupakan grafik lengkap perbandingan nilai naïve bayes dengan dan tanpa fitur chi-square. Berdasarkan grafik di atas dapat dijelaskan bahwa hasil perbandingan kinerja Naive Bayes tanpa fitur dan Naive Bayes dengan fitur Chi-Square pada dua pembagian data training dan test (80:20 dan 70:30). Feature selection dengan Chi-Square meningkatkan performa model Naive Bayes secara signifikan di semua metrik. Pembagian data latih dan uji dengan rasio 70:30 lebih efektif dibandingkan 80:20 karena memberikan hasil evaluasi yang lebih tinggi. Kombinasi terbaik diperoleh pada model NB+CS dengan split 70:30, yang mencatatkan nilai tertinggi dan paling seimbang di semua metrik evaluasi. Model Naive Bayes tanpa fitur seleksi menunjukkan hasil evaluasi yang jauh lebih rendah, memperkuat pentingnya tahap seleksi fitur dalam proses klasifikasi.

#### 4. KESIMPULAN

Evaluasi dilakukan terhadap performa algoritma klasifikasi Naive Bayes (NB) dengan dan tanpa metode feature selection menggunakan Chi-Square (CS). Eksperimen dilakukan dengan dua skenario pembagian data latih dan data uji, yaitu 80:20 dan 70:30. Tujuan dari evaluasi ini adalah untuk mengetahui pengaruh feature selection dan rasio pembagian data terhadap performa model dalam hal accuracy, precision, recall, dan F1-score. Model yang menggunakan metode Chi-Square (NB+CS) secara konsisten menunjukkan peningkatan performa pada seluruh metrik dibandingkan model tanpa Chi-Square (NB). Hal ini menunjukkan bahwa Chi-Square berhasil membantu dalam memilih fitur-fitur yang lebih relevan, sehingga meningkatkan kualitas prediksi model. Baik pada model NB maupun NB+CS, rasio 70:30 menghasilkan performa yang lebih baik dibandingkan 80:20. Hal ini disebabkan oleh lebih banyaknya data latih pada skenario 70:30, yang membantu model dalam proses pembelajaran dan generalisasi terhadap data uji. Feature selection dengan Chi-Square meningkatkan performa model Naive Bayes secara signifikan di semua metrik. Sebelum menggunakan Chi-Square hasil akurasi adalah 80,66% dan setelah menggunakan Chi-Square akurasi meningkat menjadi 86,22%. Terjadi peningkatan sebesar 5.56%. Kombinasi NB + CS dengan split 70:30 memberikan hasil terbaik, sehingga menjadi pilihan yang optimal. Penelitian selanjutnya adalah dengan menguji analisis sentimen berbahasa asing dengan machine learning lainnya seperti C4.5, XGBoost dan lain-lain

#### REFERENCES

- [1] F. Panjaitan *et al.*, “Studi Komparatif Algoritma Machine Learning Pada Analisis Sentimen Media Sosial,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 2, pp. 3145–3152, 2025.
- [2] J. Jasmir, E. Rohaini, M. R. Pahlevi, D. Sintong, and P. Simanjuntak, “Word Embedding Feature for Improvement Machine Learning Performance in Sentiment Analysis Disney Plus Hotstar Comments,” *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 10, no. 2, pp. 290–301, 2024, doi: 10.26555/jiteki.v10i2.28799.
- [3] E. S. Alamoudi and N. S. Alghamdi, “Sentiment classification and aspect-based sentiment analysis on yelp reviews using deep learning and word embeddings,” *J. Decis. Syst.*, vol. 30, no. 2–3, pp. 259–281, 2021, doi: 10.1080/12460125.2020.1864106.
- [4] I. A. Darmawan, M. F. Randy, I. Yunianto, M. M. Mutoffar, and M. T. P. Salis, “Penerapan Data Mining Menggunakan Algoritma Apriori Untuk Menentukan Pola Golongan Penyandang Masalah Kesejahteraan Sosial,” *Sebatik*, vol. 26, no. 1, pp. 223–230, 2022, doi: 10.46984/sebatik.v26i1.1622.
- [5] O. A. Alcántara Francia, M. Nunez-del-Prado, and H. Alatrística-Salas, “Survey of Text Mining Techniques Applied to Judicial Decisions Prediction,” *Appl. Sci.*, vol. 12, no. 20, 2022, doi: 10.3390/app122010200.
- [6] J. Guerreiro and P. Rita, “How to predict explicit recommendations in online reviews using text mining and sentiment analysis,” *J. Hosp. Tour. Manag.*, vol. 43, no. July, pp. 269–272, 2020, doi: 10.1016/j.jhtm.2019.07.001.
- [7] K. N. Reddy and D. B. I. Reddy, “Restaurant Review Classification Using Naives Bayes Model,” *J. Univ. Shanghai Sci. Technol.*, vol. 23, no. 08, pp. 646–656, 2021, doi: 10.51201/jusst/21/08443.
- [8] A. Abdulllah, S. Putri, and A. Alkadri, “Classification of Fetal Health Using the K-Nearest Neighbor Method and the Relief Feature Selection Method,” *J. Artif. Intell. Eng. Appl.*, vol. 4, no. 2, pp. 2–5, 2025.
- [9] V. Junita and F. A. Bachtar, “Klasifikasi Aktivitas Manusia menggunakan Algoritme Decision Tree C4.5 dan Information Gain untuk Seleksi Fitur,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 10, pp. 9426–9433, 2020, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/6446>
- [10] H. A. R. Harpizon, R. Kurniawan, Iwan Iskandar, R. Salambue, E. Budianita, and F. Syafria, “Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naïve Bayes,” ... *Di YouTube ...*, vol. 5, no. 1, pp. 131–140, 2022.
- [11] E. Indrayuni, “Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes,” *J. Khatulistiwa Inform.*, vol. 7, no. 1, pp. 29–36, 2019, doi: 10.31294/jki.v7i1.1.
- [12] R. Sari, “Analisis Sentimen Review Restoran menggunakan Algoritma Naive Bayes berbasis Particle Swarm Optimization,” *J. Inform.*, vol. 6, no. 1, pp. 23–28, 2019, doi: 10.31311/ji.v6i1.4695.
- [13] R. Azizah Arilya, Y. Azhar, and D. Rizki Chandranegara, “Sentiment Analysis on Work from Home Policy Using Naïve Bayes Method and Particle Swarm Optimization,” *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 7, no. 3, p. 433, 2021, doi: <https://doi.org/10.26555/jiteki.v7i3.22080>.
- [14] F. Septianingrum and A. S. Y. Irawan, “Metode Seleksi Fitur Untuk Klasifikasi Sentimen Menggunakan Algoritma Naive Bayes: Sebuah Literature Review,” *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 799, 2021, doi: 10.30865/mib.v5i3.2983.
- [15] O. Irnawati and K. Solecha, “Analisis Sentimen Ulasan Aplikasi Flip Menggunakan Naïve Bayes dengan Seleksi Fitur PSO,” *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 4, no. 02, pp. 189–199, 2022, doi: 10.46772/intech.v4i02.868.
- [16] E. Jasmir, Jasmir, Rasywir, H. Yani, and A. Nugroho, “Comparison of word embedding features using deep learning in sentiment analysis,” *TELKOMNIKA Telecommun. Comput. Electron. Control*, vol. 23, no. 2, pp. 416–425, 2025, doi: 10.12928/TELKOMNIKA.v23i2.26223.



- [17] V. N. February, M. A. Daniel, S. Chong, L. Chong, and K. Wee, "Optimising Phishing Detection : A Comparative Analysis of Machine Learning Methods with Feature Selection," *Journal of Informatics and Web Engineering*, vol. 4, no. 1, 2025.
- [18] S. Bahassine, A. Madani, M. Al-Sarem, and M. Kissi, "Feature selection using an improved Chi-square for Arabic text classification," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 32, no. 2, pp. 225–231, 2020, doi: 10.1016/j.jksuci.2018.05.010.
- [19] M. B. Hamzah, "Classification of Movie Review Sentiment Analysis Using Chi-Square and Multinomial Naïve Bayes with Adaptive Boosting," *J. Adv. Inf. Syst. Technol.*, vol. 3, no. 1, pp. 67–74, 2021, doi: 10.15294/jaist.v3i1.49098.
- [20] A. Falasari and M. A. Muslim, "Optimize Naïve Bayes Classifier Using Chi Square and Term Frequency Inverse Document Frequency For Amazon Review Sentiment Analysis," *J. Soft Comput. Explor.*, vol. 3, no. 1, pp. 31–36, 2022, doi: 10.52465/josce.v3i1.68.
- [21] N. Wijaya, "Evaluation of Naïve Bayes and Chi-Square performance for Classification of Occupancy House," *Int. J. Informatics Comput.*, vol. 1, no. 2, p. 46, 2020, doi: 10.35842/ijicom.v1i2.20.