

# Perbandingan Pelabelan Data dalam Analisis Sentimen Kurikulum Proyek di platform TikTok: Pendekatan Naïve Bayes

Anissya Agsani Pratiwi<sup>1</sup>, Mia Kamayani<sup>2</sup>

<sup>1,2</sup>Teknik Informatika, Fakultas Teknologi Industri dan Informatika

<sup>1,2</sup>Universitas Muhammadiyah Prof. Dr. HAMKA

<sup>1,2</sup>Jakarta, Indonesia

e-mail: <sup>1</sup>anisaagsani@gmail.com, <sup>2</sup>mia.kamayani@uhamka.ac.id

Diajukan: 15 Januari 2024; Direvisi: 6 Maret 2024; Diterima: 7 Maret 2024

## Abstrak

Penelitian ini fokus pada analisis sentimen mahasiswa terhadap perubahan kurikulum berbasis proyek di tingkat pendidikan tinggi yang menghilangkan kewajiban skripsi, Data sentimen diekspresikan melalui platform media sosial TikTok, dan algoritma Naïve Bayes digunakan untuk mengklasifikasikan sentimen menjadi positif atau negatif. Proses penelitian mencakup pengambilan data, pembersihan data, preprocessing data, pelabelan data, hingga klasifikasi menggunakan algoritma Naïve Bayes. Penelitian ini melibatkan dua tahap pelabelan dalam 913 data: pelabelan pertama manual menghasilkan 510 sentimen positif dan 403 negatif, sementara pelabelan kedua otomatis dengan RapidMiner menghasilkan 415 sentimen positif dan 498 negatif. Beberapa mahasiswa memberikan ulasan positif menganggap hal ini sebagai langkah inovatif untuk persiapan di dunia kerja. Meskipun beberapa merasa khawatir dengan tingkat kesulitan yang lebih tinggi. Hasil penelitian menunjukkan mayoritas tanggapan positif terhadap kurikulum berbasis proyek, dengan nilai pelabelan manual mencapai accuracy 93.98%, precision 100%, recall 87.99%. Sedangkan pelabelan otomatis dengan Rapidminer memperoleh nilai accuracy 70.41%, precision 80.15%, recall 69.96%.

**Kata kunci:** Analisis Sentimen, Algoritma Naïve Bayes, Skripsi, Proyek, TikTok.

## Abstract

This research focuses on the sentiment analysis of students towards the project-based curriculum changes in higher education that eliminate the thesis requirement. Sentiment data are expressed through the TikTok sosial media platform, and the Naïve Bayes algorithm is employed to classify sentiments as positive or negative. The research process includes data collection, data cleaning, data preprocessing, data labeling, and classification using the Naïve Bayes algorithm. The study involves two labeling stages with 913 data points: the first manual labeling produces 510 positive sentiments and 403 negative sentiments, while the second automatic labeling with RapidMiner results in 415 positive sentiments and 498 negative sentiments. Some students provide positive reviews, considering it an innovative step for preparation in the professional world, although some express concerns about the higher level of difficulty. The research findings indicate a majority of positive responses to the project-based curriculum, with manual labeling achieving an accuracy of 93.98%, precision of 100%, and recall of 89.22%. Meanwhile, automatic labeling with RapidMiner obtains an accuracy of 70.41%, precision of 71.17%, and recall of 59.73%.

**Keywords:** Sentiment Analysis, Algorithm Naïve Bayes, Thesis, Project, TikTok.

## 1. Pendahuluan

Perguruan tinggi adalah puncak dari tingkat Pendidikan formal dimana mahasiswa mempersiapkan dirinya sebelum memasuki dunia nyata. Mahasiswa akan diuji atas kemampuan dan kesiapannya melalui penyelesaian tugas sebagai persyaratan untuk meraih gelar sarjana berupa skripsi. Penulisan skripsi merupakan wujud nyata upaya mahasiswa menerapkan ilmu yang telah dipelajarinya selama di perguruan tinggi sebagai tanda dari integritas mereka [1].

Perkembangan ilmu pengetahuan di era modern yang mengalami kemajuan sangat pesat, ilmu pengetahuan dan teknologi ini menghasilkan inovasi dan penemuan baru setiap harinya. Baru-baru ini muncul pemberitahuan mengenai perubahan mahasiswa S1/D4 tidak diwajibkan skripsi sebagai syarat

kelulusan tetapi prodi mahasiswa harus menerapkan kurikulum berbasis proyek/lainnya, Sesuai dengan aturan yang telah ditegaskan oleh Menteri Pendidikan Kebudayaan, Riset, dan Teknologi (Permendikbudristek). Namun mahasiswa memiliki pandangan yang berbeda terhadap perubahan kurikulum ini, sebagian mahasiswa mungkin merasa senang dan mendukung perubahan kurikulum berbasis proyek, menjadikan peluang untuk mengembangkan dan mengaplikasikan pengetahuan secara langsung dalam situasi nyata, sementara mahasiswa yang lain mungkin merasa terbebani dengan perubahan ini merasa sulit beradaptasi dengan perubahan. Beberapa mahasiswa mungkin membuat sentimen pribadi mereka di media sosial dengan memberikan penilaian terhadap kurikulum berbasis proyek.

Media sosial adalah Platform yang menyediakan dukungan bagi pengguna untuk bekerjasama dalam menciptakan konten yang dihasilkan oleh pengguna. Sudah ada banyak platform media sosial seperti Instagram, TikTok, Facebook, dan Twitter [2]. TikTok menjadi salah satu media sosial platform yang populer saat ini yang dibuat oleh Zhang Yiming pada bulan September 2016 [3]. Pengguna Platform media sosial TikTok semakin meningkat dan menjadi favorit di kalangan masyarakat. Tidak hanya digunakan oleh kalangan umum, tetapi juga dimanfaatkan oleh para pelajar sebagai alat untuk mengekspresikan diri, berbagi informasi, dan melibatkan diri dalam berbagai kegiatan [4]. Terdapat banyak fitur di dalam aplikasi TikTok seperti video, filter, dan kolom komentar. Kolom komentar ini menjadi tempat interaksi pengguna untuk berbagi pendapat, memberikan feedback, atau berkomunikasi dengan pengguna lain mengenai konten yang mereka tonton.

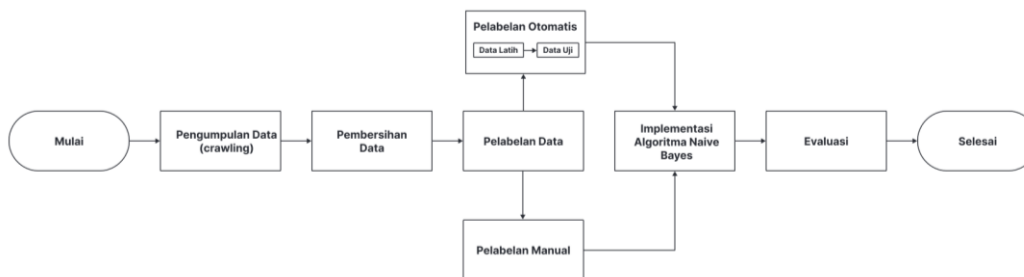
Dalam penelitian ini banyaknya opini mahasiswa terhadap perubahan kurikulum berbasis proyek yang mereka tuangkan lewat media sosial melalui komentar pada TikTok maka, pengumpulan hasil data dari TikTok melalui cara crawling data. Hasil dari proses crawling data adalah kumpulan dokumen atau dataset yang dapat digunakan untuk analisis sentiment [5]. Analisis sentimen melibatkan pengelompokan polaritas dalam suatu kalimat, dengan tujuan menentukan apakah kalimat atau opini tersebut mencerminkan aspek positif, negatif, atau bersifat netral [6]. Analisis sentimen atau opinion mining merujuk dalam ranah yang mencakup dalam pemrosesan bahasa alami, komputasi linguistik, dan text mining. Tujuan utamanya untuk menganalisis pendapat, sentimen, penilaian, sikap, serta emosi oleh seseorang, baik itu dalam berbicara maupun tulisan, terkait suatu topik, produk, layanan, organisasi, individu, atau aktivitas khusus [7]. Peneliti memanfaatkan metode Naïve Bayes sebagai algoritma yang sering digunakan dalam klasifikasi dokumen. Algoritma ini mempunyai nilai yang berpotensi baik dalam klasifikasi kecepatan dan akurasi yang tinggi [8]. Tujuan penggunaan algoritma ini untuk mencapai tingkat akurasi yang tinggi dalam menentukan apakah opini pada data TikTok bersifat negatif atau positif [9]. *RapidMiner* memberikan dukungan untuk mengambil keputusan yang lebih baik, dan memberikan solusi yang efektif dengan menerapkan berbagai algoritma serta prinsip data mining, analisis prediksi, *text mining*, dan *data mining* [10].

Pada penelitian sebelumnya yang membahas tentang analisis sentiment dengan judul penelitian "Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter," hasil analisis sentiment menandakan hasil algoritma klasifikasi *Multinomial Naïve Bayes* berhasil mencapai tingkat *accuracy* sebesar 74 %, *precision* sebesar 74 % dan *recall* 74 %. Hal ini menunjukkan bahwa kinerja algoritma Naïve Bayes cukup efektif pada saat mengolah data yang diambil dari opini Twitter [11]. Pada penelitian berikutnya yaitu penelitian yang sama oleh [12] berjudul "Naïve Bayes dan Wordcloud untuk Analisis Sentimen Wisata Halal Pulau Lombok" Penelitian ini membahas analisis sentimen pariwisata halal di Pulau Lombok yang mencapai *accuracy* sebesar 74,75%, *recall* sebesar 75,45%, dan *precision* sebesar 97,84%. Visualisasi Wordcloud menunjukkan kata-kata yang paling kerap muncul, seperti "indah," "wisata," "pantai," "alam," "gunung," dan "masjid".

Dengan hasil penelitian dilakukan diharapkan bahwa analisis sentimen yang diperoleh bisa bermanfaat dalam menilai atau mengevaluasi efektivitas Analisis Sentimen Kurikulum Berbasis Proyek dengan pendekatan Algoritma Naïve Bayes di Platform TikTok. Bertujuan untuk memberi wawasan penting tentang persepsi dan respons masyarakat terhadap kurikulum berbasis proyek.

## 2. Metode Penelitian

Dalam penelitian ini, memanfaatkan algoritma Naïve Bayes yang melibatkan beberapa tahapan dalam melakukan klasifikasi mengenai ulasan perubahan kurikulum berbasis proyek. Rangkaian langkah yang akan dijalankan dalam penelitian terdapat pada diagram yang disajikan di bawah ini. Diagram ini berperan sebagai alat bantu untuk memberikan pandangan yang lebih terinci dan terstruktur terkait dengan proses dan langkah-langkah yang akan diambil pada penelitian, memudahkan pemahaman serta pemetaan seluruh proses yang akan dijalankan dalam penelitian ini dapat dilihat seperti yang tergambar dibawah ini:



Gambar 1. Metode penelitian.

**2.1. Pengumpulan Data (Crawling Data)**

Pada tahap ini merupakan langkah awal dalam proses pengumpulan data opini mahasiswa terkait perubahan kurikulum berbasis proyek. Proses pengumpulan data dimulai dengan memanfaatkan platform TikTok, pengumpulan data dilaksanakan pada tanggal 2 November 2023, di mana langkah awalnya adalah melakukan login ke akun TikTok melalui situs web. Setelah berhasil login, selanjutnya adalah melakukan pencarian mengenai subjek yang akan diteliti. Kemudian pada halaman web TikTok melakukan scraping data komentar yang ada pada salah satu konten TikTok dengan menggunakan fitur inspect. Setelah itu masukan script javascript untuk penarikan data. Data yang berhasil diambil, berupa tabel akan disimpan dalam format file CSV. Selanjutnya data tersebut akan di proses melalui Rapid Miner untuk analisis lebih lanjut. Terlihat pada gambar 2 proses crawling data hingga menghasilkan file csv:



Gambar 2. Proses Crawling Data.

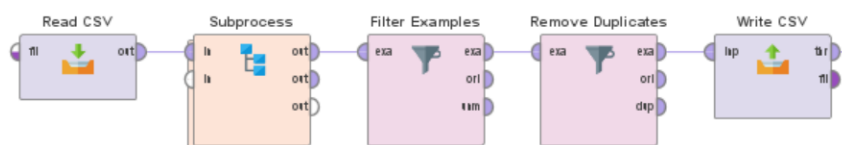
Setelah melakukan crawling data, diperoleh sebanyak 1243 data yang mencakup berbagai informasi penting terkait topik yang diteliti. Proses ini dilakukan dengan menarik data melalui komentar pada salah satu akun "Vokasia" yang membahas tentang perubahan kurikulum proyek. Berikut hasil file csv yang didapatkan pada proses crawling data terlihat pada gambar 3 dibawah ini:

1	nana	onaadddd	https://w/betul sidang itu yg menakutkan sidangnya makanya aku ga daftar" ??
2	HEROISSA	zyns_	https://w/semangat kk semoga cpet lulus??
3	nana	onaadddd	https://w/aminn doain bulan depan sidang yakk
4	winternig	genetikjab	https://w/mending skripsi tapi dospem nya di sadarkan biar ga menghambat jalan mahasiswa :")
5	samsudin	_ygzklonic	https://w/susah kak karakter sifat orang beda?
6	KurisSto	kuristo_	https://w/untung dospem gw supportif, tiap ketemu selalu ditanya progres ama disuruh bimbingan??
7	aisyahhh	m1kteeaa	https://w/susah cuyy
8	Tiara Andi	tiaaraand	https://w/apapun tugas akhirnyaa semoga aku & temen-temen selalu dimudahkan yaaa ??
9	e??	cakebys4s	https://w/aamiinn @ayoitspika @naaa
10	ini pikaa?	ayoitspika	https://w/aamiin ya rab
11	?????????	anataacaa	https://w/aamiinn @carousel?? @sukuluppa @kuyaaaa @anin
12	BaseOnYo	baseonyoi	https://w/akhirnya, kan lebih enak kalo langsung proyek
13	BaseOnYo	baseonyoi	https://w/tapi tetap ga guna sih, kebanyakan univ ga bakal mudah bermigrasi, jadi pasti bakal tetap mau gunain skripsi
14	alifahar40	alifahar40	https://w/ya bener nih
15	jevy	gg88098	https://w/tdi bu ura lwt di fyp ku, udg ada univ, jurusan komunikasi klw ga salah yg adaptasi proyek besar sebagai pengganti skripsi.
16	Nabilah	bylaa05	https://w/PAK ANGAKATAN 2020 SIAP MENJADI PERCOBAAN LAGI PAK KAGA PAS PENGHAPUSAN UN??
17	Panitia Ha	sidu21x29	https://w/lama" gw kasian juga liat ak20??
18	andre	theriskian	https://w/merasa terpengal??

Gambar 3. Hasil Crawling Data.

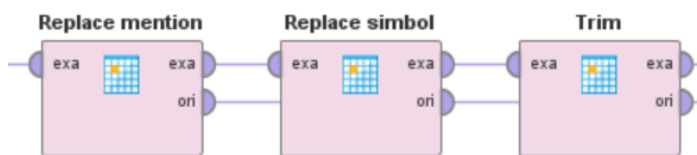
**2.2. Pembersihan Data**

Setelah melakukan *crawling* data proses selanjutnya melakukan tahap awal dalam pemrosesan data yaitu *preprocessing* atau pembersihan data tujuannya untuk menghilangkan informasi yang tidak berhubungan, melakukan normalisasi serta menyiapkan data agar dapat diolah lebih lanjut [13].



Gambar 4. Proses Pembersihan Data.

Dalam proses pembersihan data yang bertujuan untuk menghilangkan simbol – simbol dan kalimat yang tidak bermakna agar kalimat menjadi bersih maka, penelitian ini menggunakan operator yang berada pada bagian *subprocess* yaitu operator *Replace* digunakan untuk menghilangkan *mention username@*, simbol – simbol, dan tanda baca.



Gambar 5. Proses Replace.

Tabel 1. Hasil Pembersihan.

Sebelum	Sesudah
ayo pak d sah kan bkin tugas proyek lebih kreatif, kalau skripsi kdg dosen nya yg mmprsulit pak?????	ayo pak d sah kan bkin tugas proyek lebih kreatif, kalau skripsi kdg dosen nya yg mmprsulit pak
@mele @Vina @Okta Fatah @windi bagus nggak sih “kita di berpeluang buat kuliah sambil kerja, memperbanyak sistem kerja lpgan dll.	bagus nggak sih kita di berpeluang buat kuliah sambil kerja memperbanyak sistem kerja lpgan dll
iya kaa, aku maba jadi tkut anjirr?????	iya kaa aku maba jadi tkut anjirr
Akuntansi proyek ap yh? ngitung duit makin ghoib kh??	Akuntansi proyek ap yh ngitung duit makin ghoib kh

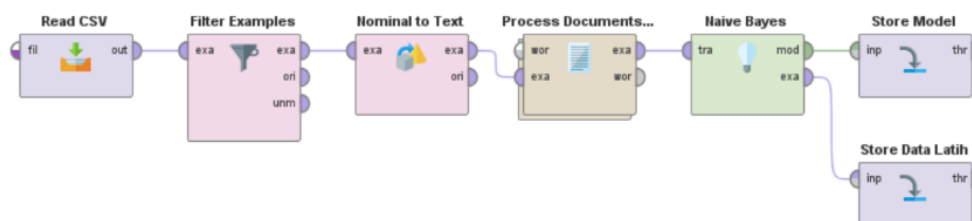
**2.3. Pelabelan Data**

Dari data yang diperoleh selama proses crawling berjumlah 1.243 data, dan setelah dataset menjalani tahap pembersihan dengan menghilangkan simbol-simbol yang tidak relevan, jumlah data yang tersisa adalah sebanyak 913 data. Proses selanjutnya adalah melakukan pelabelan data untuk menilai apakah sentimen tersebut mengarah positif atau negatif dengan membaginya menjadi dua kelompok. Dalam konteks ini, Peneliti akan melakukan perbandingan pada pelabelan data tersebut. Pertama, dataset akan dilabeli secara manual dengan jumlah 913 data. Pelabelan secara manual ini dilakukan oleh dua orang Mahasiswa Universitas Muhammadiyah Prof. Dr. Hamka dengan Program Studi Teknik Informatika. Selanjutnya peneliti akan menerapkan pelabelan secara otomatis dengan menggunakan tools RapidMiner pada dataset yang sama namun untuk melakukan pelabelan pada tahap otomatis ini diperlukan 100 data latih yang akan diberi label secara manual untuk melatih algoritma Naïve Bayes agar mampu melakukan pelabelan secara otomatis yang berjumlah 813 data. Tujuan dilakukannya pelabelan dengan dua cara ini adalah untuk membandingkan dan mengevaluasi hasil *accuracy* pelabelan dari kedua tahap berikut. Pada tabel 2 berikut proses pelabelan teks yang sudah memiliki sentimen positif dan negatif secara manual:

Tabel 2. Pelabelan Mandiri Data Latih.

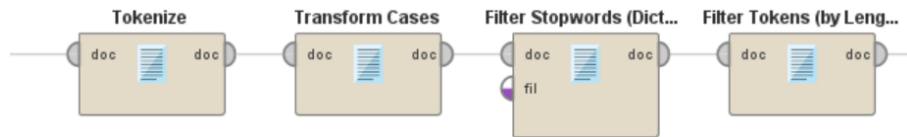
Variable	Sentimen
ayo pak d sah kan bkin tugas proyek lebih kreatif, kalau skripsi kdg dosen nya yg mmprsulit pak	Positif
bagus nggak sih kita di berpeluang buat kuliah sambil kerja memperbanyak sistem kerja lpgan dll	Positif
iya kaa aku maba jadi tkut anjirr	Negatif
Akuntansi proyek ap yh ngitung duit makin ghoib kh	Negatif

Setelah menyelesaikan Langkah pelabelan mandiri, Langkah selanjutnya adalah melakukan analisis sentiment secara otomatis menggunakan RapidMiner pada data yang masih belum memiliki label 813 data uji.



Gambar 6. Model Pemberian Sentimen Analisis Dengan Naïve Bayes.

Pada gambar 6 ini, model klasifikasi digunakan untuk membedakan apakah sentimen pada data yang belum diberi label secara mandiri memiliki nilai positif atau negatif. Setelah model Naïve Bayes ini berhasil mengidentifikasi polaritas sentimen, peneliti akan menyimpannya ke dalam operator *store*. Setelah itu operator *store* model dan *store* data latih akan digunakan untuk proses selanjutnya. Pada tahap ini terdapat operator *Process Document FromData* yang didalam operator tersebut melakukan beberapa tahapan proses *Tokenize*, *Transform Cases*, *Stopword Removal*, dan juga *Filter Tokens*.



Gambar 7. *Tokenization, Transform Cases, Filter Stopword, Filter Tokens.*

**2.3.1. Tokenization**

Proses untuk memisahkan satu kata dengan kata lainnya dari suatu kalimat. Setiap kata akan terpisah menjadi satu atribut tersendiri [14]. Proses tokenisasi ini mempermudah analisis teks dan pengolahan bahasa alami, karena teks yang telah dipecah menjadi token-token dapat diolah secara lebih terstruktur.

**2.3.2. Transform Cases**

Operator yang sering digunakan untuk mengubah huruf kapital atau huruf kecil. Sedangkan pada penelitian ini *transform cases* digunakan pada parameters *lower case* yang dapat mengubah huruf besar menjadi huruf kecil.

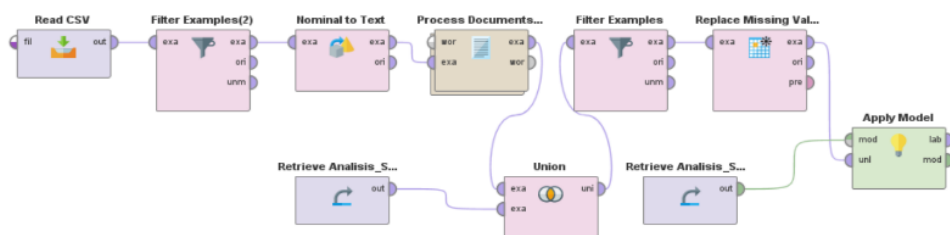
**2.3.3. Filter Stopword**

Fungsi dari operator ini untuk menghilangkan kata-kata yang bukan mempunyai makna relevan dalam analisis sentimen seperti: yang, di, adalah, dan lain-lain [15]. Proses ini melibatkan pemilihan kata-kata yang tidak memiliki makna atau tidak baku dalam suatu tweet, sehingga kata-kata tersebut tidak digunakan, dengan tujuan membuat tweet menjadi lebih bersih dan teratur [16].

**2.3.4. Filter Token by length**

Proses operator ini berfungsi untuk menyaring jumlah karakter minimal dan maksimal dari kata yang diambil dalam penelitian, sehingga mengurangi kata-kata yang kurang relevan.

Tahap selanjutnya model sentimen yang telah dibuat dengan metode Naïve Bayes di atas akan diterapkan pada data yang belum memiliki label sentimen. Dan akan diproses pelabelan secara otomatis menggunakan *RapidMiner*.



Gambar 8. Pemberian Label Otomatis.

Pada data latih dan data uji memiliki atribut yang berbeda, oleh karena itu, perlu menyamakan atribut pada kedua set data tersebut dengan melakukan penggabungan (*union*) dari kedua dataset dan melakukan proses *filter examples* untuk menghilangkan data latih yang tidak sesuai, selanjutnya ketika peneliti ingin mengintegrasikan data yang memiliki perbedaan atribut, terdapat atribut yang tidak lengkap atau kosong. oleh karena itu diperlukan operator menggantikan nilai yang hilang (*missing values*). Setelah mengikuti proses pada gambar 8 pelabelan secara otomatis menggunakan *RapidMiner* maka akan menghasilkan data yang sudah diberi label dan proses pelabelan telah selesai.

prediction(Sentimen)	text
Negatif	angkatannnnnnn kelinci percobaan
Negatif	skripsi ketentuan revisian bimbingan dibenahi mempersulit mahasiswa
Positif	semoga dilancarkan semangat
Positif	inifitly silaapajaaaa allah semoga beneran ilang
Negatif	jreng jreng
Negatif	perpajakan identifikasi pajak terkorupsi
Positif	skripsi kampus tolong gausah pake sidang
Positif	apapun ketentuannya semoga lulus permudah jalannya
Positif	bikin proyek susah
Negatif	disaat lulus
Negatif	nanggung
Negatif	dinda bener wkwk
Positif	hahaha semoga semoga

Gambar 9. Hasil Label Otomatis.

### 2.4. Penerapan Algoritma

Dataset yang telah melewati tahap pembersihan data dan pembagian pelabelan pada kedua tahap pelabelan, yaitu pelabelan manual dan pelabelan otomatis menggunakan *tools RapidMiner* akan diproses melalui perhitungan algoritma Naïve Bayes. Salah satu algoritma yang umum diterapkan adalah Algoritma Naïve Bayes untuk melakukan klasifikasi menentukan label pada ulasan sentimen, dengan tujuan mengidentifikasi apakah ulasan tersebut mengandung makna positif atau negatif [17].

### 2.5. Pengujian dan Evaluasi

Setelah melalui proses pembersihan data dan pelabelan menggunakan dua pendekatan, langkah berikutnya adalah menguji akurasi data pada klasifikasi melalui pemanfaatan metode Naïve Bayes. *Accuracy*, *precision*, dan *recall* dihitung dengan menggunakan teknik *confusion matrix*. Teknik *confusion matrix* terdiri dari : *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Nilai-nilai tersebut dapat dihitung dengan memperhatikan *confusion matrix* yang terdapat pada tabel di bawah ini.

Tabel 3. Confusion Matrix.

Actual	Predicted	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Keterangan:

- TP = *True Positive* terprediksi positif benar
- FP = *False Positive* terprediksi salah positif
- TN = *True Negative* terprediksi benar negatif
- FN = *False Negative* terprediksi salah negatif

Akurasi memberikan pemahaman tentang seberapa baik kinerja model berhasil dalam memprediksi data dengan tepat secara keseluruhan. Perhitungannya melibatkan pengelompokan jumlah prediksi yang benar dibandingkan dengan total jumlah sampel [18].

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \tag{1}$$

Precision adalah metrik yang mengukur sejauh mana model pengklasifikasi berhasil dalam memprediksi dengan benar jumlah data positif dari seluruh data yang diprediksi sebagai kelas positif [19].

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

Recall adalah metrik yang mengukur sejauh mana sistem berhasil mengukur kemampuan model yang relevan untuk mengidentifikasi secara benar semua contoh positif dalam kumpulan data [20].

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

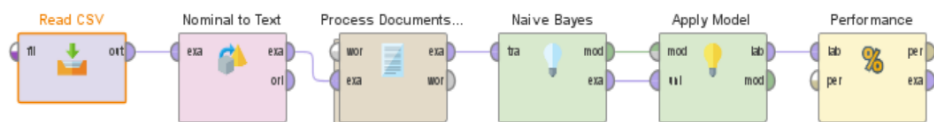
### 3. Hasil dan Pembahasan

Pada bagian ini, dilakukan penelitian dan beberapa pengujian. Proses analisis sentimen menggunakan algoritma naïve bayes pada data yang telah terkumpul. Algoritma ini digunakan untuk mengklasifikasikan sentimen dari kalimat-kalimat yang terdapat dalam penelitian. Setelah itu, dilakukan perhitungan untuk menentukan nilai akurasi, presisi, dan recall dengan menggunakan confusion matrix dalam perhitungan tersebut. Confusion matrix memungkinkan peneliti untuk mengevaluasi sejauh mana performa model dalam mengklasifikasikan sentimen positif dan negatif.

Evaluasi lebih lanjut terhadap akurasi, presisi, dan recall melalui confusion matrix juga memberikan pemahaman mendalam mengenai performa model tersebut dalam mengklasifikasikan sentimen secara tepat dan konsisten. Hasil dari pengujian ini dapat menjadi dasar untuk mengetahui keberhasilan algoritma yang digunakan dalam mencapai tujuan penelitian terkait analisis sentimen.

#### 3.1. Klasifikasi Algoritma Naïve Bayes

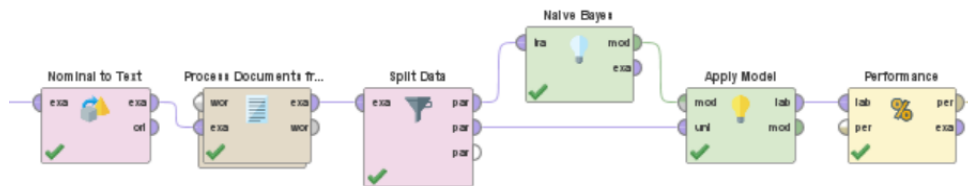
Temuan dari penelitian dibahas dan diperbandingkan menggunakan temuan dari penelitian yang diacu, kemudian pada proses ini, setelah tahap pelabelan manual pada 913, data langkah berikutnya adalah klasifikasi menggunakan algoritma naïve bayes pada *RapidMiner*. Proses ini melibatkan pemrosesan data yang telah dilabeli untuk melatih model klasifikasi.



Gambar 10. Klasifikasi Pada Pelabelan Manual.

Tujuannya adalah untuk membangun model klasifikasi yang dapat menerapkan pola dari data latih ke data yang belum pernah dilihat sebelumnya, meningkatkan tingkat ketepatan prediksi untuk proses pengambilan keputusan, dan dapat menggunakan kemampuan analisis data untuk efektif mengelola informasi.

Selain itu peneliti juga melakukan langkah klasifikasi algoritma naïve bayes pada data latih yang sudah diberi label semua data tersebut akan digunakan untuk pengujian data uji. Pengujian data uji dilakukan dengan menerapkan algoritma Naïve Bayes untuk mengidentifikasi dan memprediksi label atau kategori yang sesuai dengan setiap data uji.

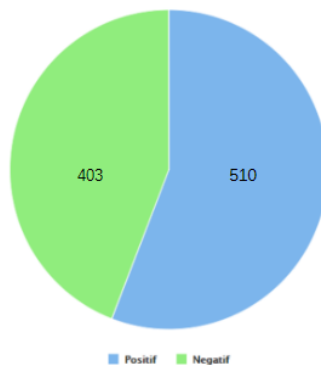


Gambar 11. Klasifikasi Pada Pelabelan Otomatis.

Pada tahap ini menggunakan Operator Split Data yang berguna untuk memisahkan dataset menjadi dua bagian yang berperan penting dalam proses pembelajaran mesin. Pada bagian parameters membaginya menjadi 20% untuk data latih dan 80% untuk data uji dengan membandingkan rasio 20:80. kemudian dihubungkan dengan algoritma naïve bayes bertujuan untuk mengklasifikasikan sentimen, dan dihubungkan juga dengan menggunakan *union* untuk menyamakan atribut pada kedua dataset. Selanjutnya evaluasi penerapan algoritma dilakukan dengan menggunakan operator *Apply Model* proses ini sebagai metode *machine learning* untuk menguji kinerja model dan memprediksi hasil berdasarkan data uji yang belum diberi label dan memanfaatkan operator *performance* dalam mengevaluasi kinerja model yang telah dibuat, membantu dalam memahami sejauh mana model tersebut efektif dalam memprediksi data.

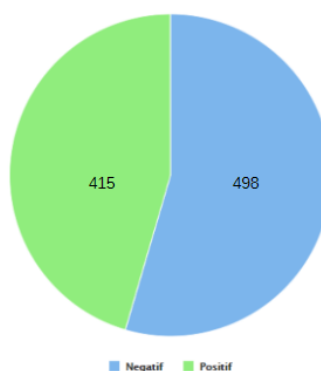
### 3.2. Visualisasi Hasil Pelabelan

Pada gambar 11 hasil penelitian ini mengungkap bahwa mayoritas tanggapan masyarakat adalah positif, dengan data positif mencapai 510 total tanggapan, sementara tanggapan negatif hanya mencapai 403. Dari data ini, dapat disarankan bahwa hasil klasifikasi pada tahap pelabelan secara manual menunjukkan adanya dominasi ulasan dengan sentimen positif dari pada sentimen negatif dalam persepsi masyarakat.



Gambar 12. Visualisasi Hasil Pelabelan Manual.

Kemudian pada gambar 13 dari hasil penelitian dibawah ini, terungkap bahwa mayoritas tanggapan masyarakat bersifat positif, dengan data positif mencapai 415 total tanggapan, sementara tanggapan negatif hanya mencapai 498. Dari hasil tersebut, dapat disimpulkan bahwa klasifikasi pada tahap pelabelan secara otomatis menggunakan *RapidMiner* menunjukkan lebih banyak ulasan sentimen negatif.



Gambar 13. Visualisasi Hasil Pelabelan Otomatis.

### 3.3. Visualisasi Wordcloud

Visualisasi adalah teknik yang digunakan untuk menyederhanakan data dan menciptakan representasi visual dari data atau informasi [21]. Seperti yang terlihat dalam gambar 12, terdapat sebuah visualisasi berupa wordcloud yang mengilustrasikan sentimen dari kalimat yang sering muncul dalam penelitian ini. Dalam wordcloud tersebut, terlihat bahwa kata-kata tertentu dipertegas dengan ukuran yang lebih besar, menciptakan efek visual di mana kata-kata yang sering muncul memiliki tampilan yang lebih dominan. Dari hasil *wordcloud* ini kata yang sering muncul yaitu: skripsi dengan total kata yang muncul berjumlah 203 kali.





Gambar 14. Hasil Wordcloud.

### 3.4. Hasil Pengujian dan Evaluasi

Selanjutnya dalam proses evaluasi, tujuannya adalah untuk memeriksa validitas hasil klasifikasi dengan menggunakan perhitungan *confusion matrix* berdasarkan penelitian terhadap sentiment Kurikulum Berbasis Proyek (Tidak Wajib Skripsi) yaitu dengan menghitung masing-masing dari dua tahap klasifikasi pada pelabelan data yang sudah dilakukan di atas.

Tabel 4. Nilai *Confusion Matrix*.

Model	TP	FP	TN	FN
Manual	403	0	455	55
Otomatis	319	79	195	137

Berikut merupakan perhitungan *confusion matrix* pada hasil pelabelan manual:

Perhitungan *Accuracy*

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} = \frac{403+455}{403+55+0+455} = \frac{858}{913} = 0,93975$$

$$Accuracy = 0,93975 \text{ atau } 93.98\%$$

Perhitungan *Precision*

$$Precision = \frac{TP}{TP+FP} = \frac{403}{403+0} = 1$$

$$Precision = 1 \text{ atau } 100\%$$

Perhitungan *Recall*

$$Recall = \frac{TP}{TP+FN} = \frac{403}{403+55} = \frac{403}{458} = 0,87991$$

$$Recall = 0,87991 \text{ atau } 87.99\%$$

Berikut merupakan perhitungan *confusion matrix* pada hasil pelabelan yang dilakukan secara otomatis menggunakan *RapidMiner*:

Perhitungan *Accuracy*

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} = \frac{319+195}{319+137+79+195} = \frac{514}{730} = 0,7041$$

$$Accuracy = 0,7041 \text{ atau } 70.41 \%$$

Perhitungan *Precision*

$$Precision = \frac{TP}{TP+FP} = \frac{319}{319+79} = \frac{319}{398} = 0,80150$$

$$Precision = 0,80150 \text{ atau } 80.15 \%$$

Perhitungan *Recall*

$$Recall = \frac{TP}{TP+FN} = \frac{319}{319+137} = \frac{319}{456} = 0,69956$$

$$Recall = 0,69956 \text{ atau } 69.96 \%$$

Tabel 5. Hasil Perhitungan Klasifikasi Naïve Bayes.

Model	Accuracy	Precision	Recall
Manual	93.98%	100%	87.99%
Otomatis	70.41%	80.15%	69.96%

Hasil temuan dari penelitian ini mengungkapkan bahwa penerapan dua metode pelabelan dalam analisis sentimen menghasilkan tingkat accuracy mencapai 93.98%, precision 100%, dan recall 87.99% pada proses pelabelan manual. Untuk memperjelas dan menemukan hasil yang lebih baik, peneliti melakukan perbandingan dengan hasil pelabelan otomatis menggunakan alat RapidMiner. Hasilnya menunjukkan bahwa pelabelan otomatis memiliki tingkat accuracy sebesar 70.41%, precision 80.15%, dan recall 69.96%. Perbandingan ini memberikan gambaran yang jelas tentang perbedaan kinerja antara pelabelan manual dan otomatis. Secara khusus, pelabelan manual mencapai hasil yang lebih tinggi dalam hal accuracy, precision dan recall, menunjukkan keunggulan model pelabelan secara manual dalam konteks analisis sentimen.

**4. Kesimpulan**

Berdasarkan analisis dan hasil penelitian terkait Sentimen Kurikulum Berbasis Proyek (Tidak Wajib Skripsi) di Platform TikTok, pengumpulan data dilaksanakan pada tanggal 2 November 2023, dengan mengumpulkan sekitar 1243 data. Setelah melalui proses pembersihan dan preprocessing data jumlah data yang tersisa adalah sebanyak 913. Pada penelitian ini menerapkan dua metode dengan melakukan pelabelan manual dan pelabelan otomatis menggunakan RapidMiner kemudian diklasifikasikan menggunakan algoritma Naïve Bayes. Pada tahap pelabelan di mana pelabelan pertama dilakukan secara manual pada 913 data menghasilkan nilai accuracy 93.98%. Kemudian pada tahap kedua dilakukan dengan pelabelan otomatis dengan data latih sebanyak 100 data dan data uji 813 data menghasilkan nilai accuracy 70.41%. Dapat diambil kesimpulan bahwa pada proses pelabelan manual ini memberikan landasan yang kuat bagi algoritma Naïve Bayes untuk melakukan klasifikasi dengan sangat tepat pada dataset yang telah diberi label secara manual. Sementara itu, pada tahap pelabelan otomatis, Meskipun accuracy 70.41 % dianggap cukup baik, penelitian ini terdapat faktor-faktor yang mempengaruhi kinerja model, seperti kualitas data latih dan data uji. Kesimpulannya ini memberikan pemahaman bahwa pelabelan otomatis dapat memberikan hasil yang baik, tetapi perlu memperhatikan dan meningkatkan kualitas data latih dan data uji untuk mencapai tingkat akurasi yang lebih tinggi. Semakin banyak data yang diberi label secara

manual, semakin tinggi tingkat akurasi Naïve Bayes dalam memberikan label otomatis. Proses ini bertujuan memberikan pelatihan kepada algoritma Naïve Bayes untuk dapat melakukan pelabelan otomatis dan mencapai tingkat akurasi yang tinggi. Saran pada penelitian berikutnya untuk dapat mempertimbangkan akurasi evaluasi atau membandingkan hasil dengan metode klasifikasi lainnya selain Naïve Bayes.

#### Daftar Pustaka

- [1] H. Wakhyudin and A. D. S. Putri, “Analisis Kecemasan Mahasiswa Dalam Menyelesaikan Skripsi,” *WASIS J. Ilm. Pendidik.*, vol. 1, no. 1, pp. 14–18, Jun. 2020, doi: 10.24176/wasis.v1i1.4707.
- [2] Y. Alkhalifi, W. Gata, A. Prasetya, and I. Budiawan, “Analisis Sentimen Penghapusan Ujian Nasional pada Twitter Menggunakan Support Vector Machine dan Naïve Bayes berbasis Particle Swarm Optimization,” *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 6, no. 2, p. 71, 2020, doi: 10.24014/coreit.v6i2.9723.
- [3] A. Purnamawati, M. N. Winarto, and M. Mailasari, “Analisis Sentimen Aplikasi TikTok menggunakan Metode BM25 dan Improved K-NN Fitur Chi-Square,” *J. Komtika (Komputasi dan Inform.)*, vol. 7, no. 1, pp. 97–105, 2023, doi: 10.31603/komtika.v7i1.8938.
- [4] M. Mariati, “Analisis Dampak Media Sosial Tik-Tok Terhadap Rendahnya Hasil Belajar Afektif Siswa,” *J. Pendidik. dan Media Pembelajaran*, vol. 2, no. 1, pp. 38–44, 2023, doi: 10.59584/jundikma.v2i1.9.
- [5] F. Amaliah and I. K. Dwi Nuryana, “Perbandingan Akurasi Metode Lexicon Based Dan Naive Bayes Classifier Pada Analisis Sentimen Pendapat Masyarakat Terhadap Aplikasi Investasi Pada Media Twitter,” *J. Informatics Comput. Sci.*, vol. 3, no. 03, pp. 384–393, 2022, doi: 10.26740/jinacs.v3n03.p384-393.
- [6] M. I. Petiwi, A. Triayudi, and I. D. Sholihati, “Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine,” *J. Media Inform. Budidarma*, vol. 6, no. 1, p. 542, 2022, doi: 10.30865/mib.v6i1.3530.
- [7] F. N. Majid and Sulastri, “Analisa Sentimen Aplikasi Pedulilindungi Dengan Metode Nbc Dan Svm,” *J. Elektron. Dan Komput.*, vol. 16, no. 1, pp. 100–108, 2023, doi: <https://doi.org/10.51903/elkom.v16i1.1000>.
- [8] B. R. Atmadja, “Analisis Sentimen Bahasa Indonesia Pada Tempat Wisata Di Kabupaten Sukabumi Dengan Naive Bayes Classifier,” *Elkom J. Elektron. dan Komput.*, vol. 15, no. 2, pp. 371–382, 2022, doi: 10.51903/elkom.v15i2.872.
- [9] N. Legiawati, T. I. Hermanto, and Y. R. Ramadhan, “Analisis Sentimen Opini Pengguna Twitter Terhadap Perusahaan Jasa Ekspedisi Menggunakan Algoritma Naïve Bayes Berbasis PSO,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, p. 930, 2022, doi: 10.30865/jurikom.v9i4.4629.
- [10] A. R. Abdillah, F. N. Hasan, T. Informatika, F. N. Hasan, D. Mining, and A. Sentimen, “Analisis Sentimen Terhadap Kandidat Calon Presiden Berdasarkan Tweets Di Sosial Media Menggunakan Naive Bayes Classifier Sentiment Analysis of Presidential Candidates Based on Tweets on,” vol. 13, no. 1, pp. 117–130, 2024, doi: <https://doi.org/10.32664/smatika.v13i01.750>.
- [11] Yuyun, Nurul Hidayah, and Supriadi Sahibu, “Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, pp. 820–826, 2021, doi: 10.29207/resti.v5i4.3146.
- [12] Irvandi, B. Irawan, and O. Nurdiawan, “Naive Bayes Dan Wordcloud Untuk Analisis Sentimen Wisata Halal Pulau Lombok,” *INFOTECH J.*, vol. 9, no. 1, pp. 236–242, 2023, doi: 10.31949/infotech.v9i1.5322.
- [13] M. Murni, I. Riadi, and A. Fadlil, “Analisis Sentimen HateSpeech pada Pengguna Layanan Twitter dengan Metode Naïve Bayes Classifier (NBC),” *JURIKOM (Jurnal Ris. Komputer)*, vol. 10, no. 2, p. 566, 2023, doi: 10.30865/jurikom.v10i2.5984.
- [14] N. Agustina, D. H. Citra, W. Purnama, C. Nisa, and A. R. Kurnia, “Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 1, pp. 47–54, 2022, doi: 10.57152/malcom.v2i1.195.
- [15] K. Anwar, “Analisa sentimen Pengguna Instagram Di Indonesia Pada Review Smartphone Menggunakan Naive Bayes,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 4, pp. 148–155, 2022, doi: 10.30865/klik.v2i4.315.
- [16] A. D. Pratama, “Analisa Sentimen Masyarakat Terhadap Penggunaan CHATGPT Menggunakan Metode Support Vector Machine (SVM),” vol. 9, no. 1, pp. 327–338, 2023, doi: <https://doi.org/10.29100/jipi.v9i1.4285>.

- 
- [17] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, “Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN,” *J. KomtekInfo*, vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [18] Ernianti Hasibuan and Elmo Allistair Heriyanto, “Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier,” *J. Tek. dan Sci.*, vol. 1, no. 3, pp. 13–24, 2022, doi: 10.56127/jts.v1i3.434.
- [19] M. Jonathan and Y. Nataliani, “Analisis Sentimen Penilaian Masyarakat Indonesia terhadap GeNose pada Komentar Youtube Menggunakan Metode Naive Bayes,” *J. Mat. dan Apl.*, vol. 11, no. 01, 2022, doi: <https://doi.org/10.35799/dc.11.1.2022.38339>.
- [20] A. Turmudi Zy, A. Nugroho, A. Rivaldi, and I. Afriantoro, “Analisis Sentimen Terhadap Pembobolan Data pada Twitter dengan Algoritma Naive Bayes,” *J. Teknol. Inform. dan Komput.*, vol. 8, no. 2, pp. 202–213, 2022, doi: 10.37012/jtik.v8i2.1240.
- [21] Y. Nurtikasari, Syariful Alam, and Teguh Iman Hermanto, “Analisis Sentimen Opini Masyarakat Terhadap Film Pada Platform Twitter Menggunakan Algoritma Naive Bayes,” *INSOLOGI J. Sains dan Teknol.*, vol. 1, no. 4, pp. 411–423, 2022, doi: 10.55123/insologi.v1i4.770.