



# Implementasi Data Mining Analisa Sentimen Program Kartu Prakerja Menggunakan Algoritma Naïve Bayes

Agistia Yuliawati\*<sup>1</sup>, Tri Wahyudi<sup>2</sup>

<sup>1,2</sup>Prodi Teknik Informatika, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Jakarta, Indonesia

\*Email Penulis Korespondensi: [agistia@stikomcki.ac.id](mailto:agistia@stikomcki.ac.id)

## Abstrak

Program kartu prakerja di luncurkan pada tanggal 11 april 2020, dengan janji untuk meningkatkan kondisi dan perekonomian melalui pengembangan keterampilan angkatan kerja, peluncuran kartu prakerja menuai banyak tanggapan positif dan negatif. Ada banyak reaksi, mulai dari diskusi tentang sistem seleksi, materi pelatihan, dan besaran anggaran. Dengan keberadaan Program Kartu Prakerja sejak pendaftaran gelombang pertama hingga Saat ini, ada sejumlah besar pengguna Twitter di Indonesia yang berbagi pendapat dan gagasan mereka tentang program tersebut melalui Twitter. Tujuan dari studi ini adalah untuk menerapkan metode Naive Bayes untuk mengevaluasi sentiment kebijakan kartu prakerja. Proses pengumpulan dataset yaitu dengan cara mengumpulkan opini pada Twitter, lalu dilakukan tahapan preprocessing pada data yang meliputi case folding, cleansing data, tokenizing, stopword, case normalization, dan stemming, selanjutnya dengan Membantu mengurangi noise yang disebabkan oleh label yang tidak dapat diterima. Pendapat dikelompokkan menjadi dua kelompok, yaitu sentimen negatif dan sentimen positif. Metode yang digunakan untuk menganalisis opini masyarakat menggunakan data dari Twitter Naïve Bayes, karena metode ini cukup mudah dan tidak membutuhkan banyak training data. Dari hasil pengujian menunjukkan rata-rata akurasi sebesar 77,58%.

**Kata kunci**— Algoritma Naïve Bayes, Data Mining, Kartu Prakerja

## Abstract

The pre-employment card program was launched on April 11 2020. With a promise to improve conditions and the economy through developing workforce skills, the launch of the pre-employment card received many positive and negative responses. There were many reactions, starting from discussions about the selection system, training materials, and budget size. With the existence of the Pre-Employment Card Program since the first wave of registration until now, there are a large number of Twitter users in Indonesia who share their opinions and ideas about the program via Twitter. The aim of this study is to apply the Naive Bayes method to evaluate pre-employment card policy sentiment. The process of collecting the dataset is by collecting opinions on Twitter, then carrying out preprocessing stages on the data which include case folding, data cleansing, tokenizing, stopwords, case normalization, and stemming, then by helping to reduce

*noise caused by unacceptable labels. Opinions are grouped into two groups, namely negative sentiment and positive sentiment. The method used to analyze public opinion uses data from Twitter Naïve Bayes, because this method is quite easy and does not require a lot of training data. The test results show an average accuracy of 77.58%.*

**Keywords**— *Naïve Bayes Algorithm, Data Mining, Pre-Employment Cards*

## 1. PENDAHULUAN

Pengangguran merupakan suatu masalah yang menghambat suatu Negara, tingkat pengangguran dapat digunakan sebagai ukuran kondisi ekonomi Negara tersebut. Pengangguran sendiri terjadi ketika orang yang sudah memasuki usia kerja tidak dapat dipekerjakan. Pengangguran pasti akan terjadi di setiap Negara dalam bentuk sosial maupun ekonomi. Angka pengangguran di Indonesia semakin meningkat selama 5 tahun terakhir, menurut laporan Badan Pusat Statistik (BPS).

Dilihat dari data *survey* BPS menunjukkan diagram pengangguran di Indonesia dari tahun 2019 hingga 2024 berdasarkan data yang tersedia. Pada tahun 2019 mencapai 7.104,42 orang, dan mengalami kenaikan yang signifikan pada tahun 2020 menjadi 9.767,75 orang. Tahun berikutnya yaitu 2021, terjadi penurunan menjadi 9.102,05 orang. Penurunan kembali terjadi pada tahun 2022 menjadi 8.425,93 orang. Namun kembali mengalami penurunan pada tahun 2023, menjadi 7.195,00 orang. Selanjutnya mengalami penurunan kembali pada tahun 2024, menjadi 7.200,00 orang [1].

Data *survey* BPS juga menunjukkan bahwa tingkat pengangguran di Indonesia telah menurun secara signifikan sejak 2019, dengan penurunan yang paling dramatis terjadi antara 2020 dan 2021. Hal ini diduga karena lesunya pertumbuhan ekonomi dan terjadinya Covid-19 yang mengakibatkan akses masyarakat terbatas untuk beraktivitas sehingga pengangguran mengalami peningkatan serta banyaknya orang yang di PHK dalam bekerja. Sehingga Pemerintah Indonesia berusaha meningkatkan ekonomi negara dengan mempercepat penerbitan kartu prakerja [2].

Salah satu kebijakan pemerintah, yang dibuat oleh Kementerian Ketenagakerjaan, adalah kartu prakerja. Menggabungkan keterampilan, meningkatkan efisiensi dan daya saing, dan mendorong kewirausahaan melalui program insentif dan biaya pelatihan. Program kartu prakerja ini ditujukan kepada individu yang sedang mencari pekerjaan, karyawan, atau karyawan yang terkena dampak kehilangan pekerjaan, serta bisnis mikro yang mengalami penurunan daya beli atau bahkan kehilangan karyawan (PHK) [3].

Berbeda dengan model sebelumnya yang bersifat semi-bansos, pemerintah akan melanjutkan program Kartu Prakerja dengan model standar. Sejak dimulai pada 2020 hingga 2022, program ini berfungsi sebagai program bansos dan peningkatan kompetensi pekerja, sehingga menjadi program semi-bansos. fokusnya adalah untuk meningkatkan kemampuan dan produktivitas karyawan [4].

Sejak pendaftaran gelombang pertama, Pendapat tentang kebijakan kartu prakerja terus berkembang, baik pro maupun kontra. Pelatihan yang dianggap terlalu mahal dengan kualitas yang buruk dianggap sebagai pemborosan anggaran selain dari sistem seleksi yang buruk, Netizen juga mengkritik bahwa pelatihan kartu pekerja yang hanya 15 jam tidak memastikan bahwa peserta dapat menyerap ilmu yang diperoleh. Meskipun banyak perdebatan, minat masyarakat terhadap program kartu prakerja tetap tinggi, hal ini menunjukkan betapa pentingnya program ini bagi masyarakat [5].

Berdasarkan sentimen, *Twitter* telah berkembang menjadi salah satu *platform* media sosial favorit banyak masyarakat, termasuk Indonesia, untuk menyampaikan minat, pendapat, dan hobi mereka. Kami dapat membagi pro dan kontra kartu prakerja, yang sedang diperdebatkan di *Twitter*, komentar positif dan negatif, menjadi dua kategori utama. Akibatnya, karena banyaknya pendapat dan komentar yang disampaikan oleh masyarakat pengguna *Twitter* dijadikan sumber data penelitian ini [6].

Analisis sentimen dapat digunakan untuk menilai seberapa efektif kebijakan pemerintah untuk membantu masyarakat, sehingga pemerintah dapat memperbaiki kekurangan mereka. Evaluasi kebijakan bertujuan untuk menentukan faktor-faktor yang menyebabkan kebijakan tertentu gagal untuk memastikan apakah kebijakan publik telah diterapkan memiliki efek yang diinginkan. Dengan mempertimbangkan masalah saat ini, penelitian ini menggunakan metode naïve bayes untuk menganalisis sentimen terhadap kebijakan kartu prakerja. Tujuannya adalah untuk mengetahui bagaimana masyarakat Indonesia memandang kebijakan tersebut [7].

Studi sebelumnya menemukan pro dan kontra program kartu prakerja melalui analisis sentimen. Setelah menggunakan metode *Matrix Confusion*. Metode Support Vector Machine (SVM) digunakan untuk menganalisis opini masyarakat menggunakan data dari *Twitter* sosial media. Di sisi lain, metode *Confusion Matrix* digunakan untuk mengukur kinerja klasifikasi SVM. Dalam penelitian ini, perbandingan kernel linear dan RBF dilakukan. Hasilnya menunjukkan bahwa kernel linear memiliki keakuratan 99%, keakuratan 98%, dan akurasi 98,67% dan skor F1 98,34%. Berdasarkan hasil analisis, dapat disimpulkan bahwa persepsi pengguna *Twitter* terhadap program kartu prakerja selama pandemi lebih netral sebesar 98,34% [8].

*Twitter* telah menjadi salah satu *platform* media sosial yang paling populer di banyak Negara, termasuk Indonesia, untuk berbagi pendapat dan hobi mereka. Untuk penyempurnaan kebijakan, pro dan kontra kartu prakerja harus dipertimbangkan. Untuk menganalisis respon masyarakat dengan data *Twitter*, metode pengklasifikasian *Naïve Bayes* dapat digunakan untuk melakukan analisis sentimen. Dari model pengklasifikasian data *original*, *training*, dan *testing*, diperoleh persentase respons berupa sentimen negatif terkait kartu prakerja sebesar 52,87%, lebih tinggi dari persentase sentimen positif sebesar 47,13%, dan juga diperoleh nilai akurasi sebesar 91,06% dari total tweet yang dikirim [9].

## 2. METODE PENELITIAN

### 2.1 Data untuk Penelitian

Semua data penelitian ini adalah fakta dan angka. digunakan untuk menyusun informasi. Pada fase ini, data dikumpulkan melalui *Crawling Data Google Colab* Program Kartu Prakerja.

#### 2.1.1 Lokasi Penelitian

Peneliti melakukan penelitian di tempat ini. dan mendapatkan informasi tentang data yang dikumpulkan. Lokasi penelitian perlu dipilih karena sesuai dengan topik penelitian. Tempat penelitian adalah melalui *Crawling Data Google Colab* pada media sosial *Twitter*.

#### 2.1.2 Waktu untuk Penelitian

Rencana penelitian, pengajuan judul, penyusunan proposal, penyelenggaraan penelitian, seminar proposal, revisi proposal, dan penyusunan skripsi adalah semua bagian dari waktu penelitian. Proses ini akan dimulai pada Mei hingga Juni 2024.

#### 2.1.3 Metode Pengumpulan Informasi Data

Pada tahap pengumpulan data, dataset untuk Program Kartu Prakerja tahun 2023, dikumpulkan melalui *Crawling Data Google Colab*. *Dataset* ini berisi 1.371 baris data, yang menghasilkan dua kelompok data: data positif dan data negatif. Secara keseluruhan, 1.008 data yang digunakan dalam studi adalah data kualitatif. Studi ini melakukan analisis kualitatif untuk menentukan tingkat akurasi yang paling tinggi dan hasil sentimen pengguna *Twitter*, apakah itu positif atau negatif, yang dapat memengaruhi perspektif masyarakat tentang Program Kartu Prakerja. Aplikasi *Microsoft Excel* dan *Rapidminer* membantu memproses dan membuat laporan tentang data *Twitter* [10]. Aplikasi *Mendeley Desktop* membantu memasukkan referensi ke laporan penelitian. *Dataset* publik digunakan. Data kualitatif terdiri dari dua bagian, termasuk:

### 1. *Data Primer(Utama)*

Data yang diperoleh langsung dari individu yang terlibat. Metode untuk memperoleh data Primer melalui *observasi*, dengan melakukan observasi tanggapan penelitian pada *Twitter*, menimbulkan tanggapan Pro dan Kontra. *Data primer* yang dikumpulkan melalui *observasi* termasuk *Hashtags, Username, Tweet Text, Replies, Retweet.* dan *Time Stamp*.

### 2. *Data Sekunder(Tambahan)*

Jenis data tambahan yang dikirim oleh sumber disebut data sekunder, data kesekian daripada dikumpulkan dari sumber utama. Data sekunder digunakan untuk meningkatkan pengolahan data primer. Studi Pustaka dan *Textbook* adalah contoh data sekunder.

#### 2.1.4 *Atribut Data Penelitian*

Data mentah dengan 6 atribut dan keterangannya disajikan dalam Tabel 1, yang terdiri dari 6 buah fitur.

Tabel 1 Atribut *Dataset*

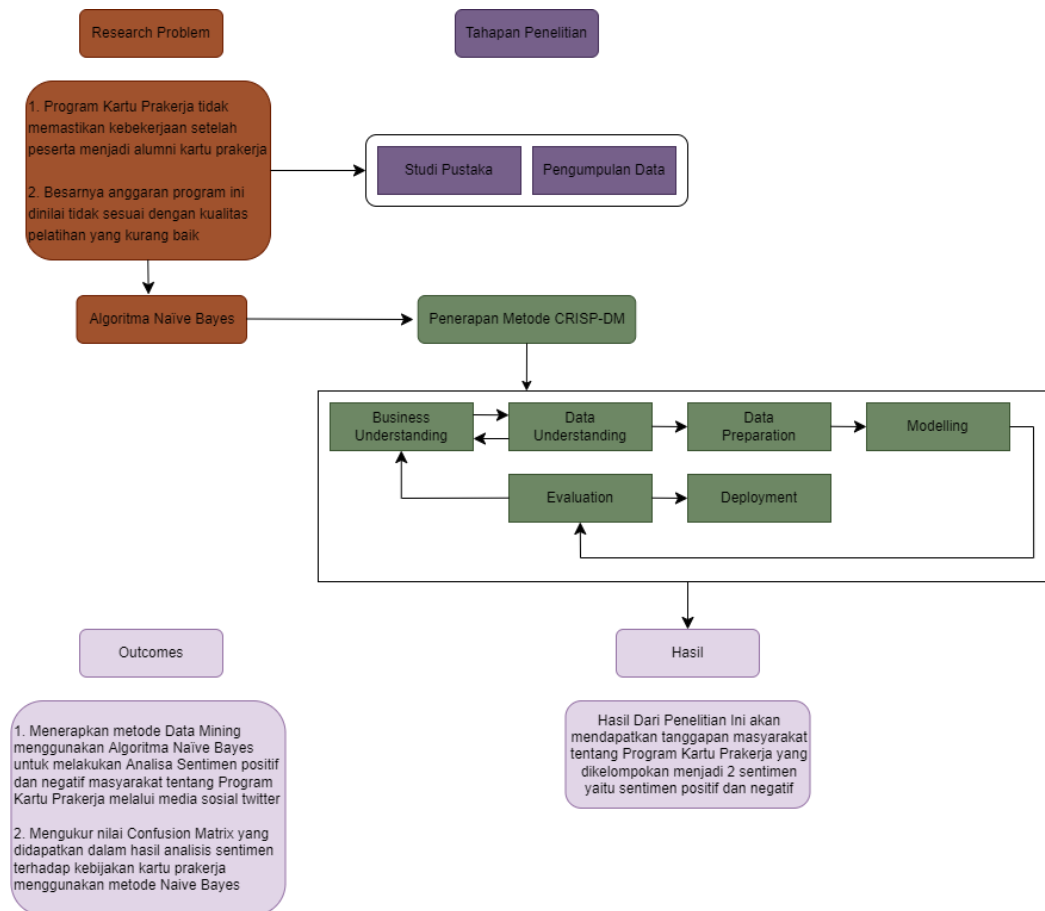
Atribut	Keterangan
<i>Hashtags</i>	Tagar untuk mengaitkan
<i>Username</i>	daftar pengguna
<i>Replies</i>	Respon terhadap pesan
<i>Time Stamp</i>	ketika tweet dikirim
<i>Retweet</i>	pembagian <i>tweet</i>
<i>Tweet Text</i>	Isi pada <i>tweet</i>

Data ini diambil dari tanggapan masyarakat mengenai Program Kartu Prakerja pada sosial media *Twitter*. Setelah dikumpulkan, data harus diubah dari data mentah menjadi data yang telah diolah untuk proses selanjutnya, yang mencakup pendeskripsian data, evaluasi pemilihan data, dan pemilihan atribut. Kemudian dilakukan pengolahan data menggunakan algoritma *Naïve Bayes* menggunakan *Rapidminer*.

#### 2.1.5 *Data Jumlah*

Data studi ini diperoleh dari *Crawling Data Google Colab* Program Kartu Prakerja melalui media sosial twitter berjumlah 1.371 data.

## 2.2 Penerapan Metodologi



Gambar 1 Tahap Penerapan Metodologi

Studi ini menganalisis perasaan masyarakat dengan menggunakan pendekatan *Naive Bayes*, tentang tanggapan Program Kartu Prakerja di *Twitter*. Mengacu pada Gambar 1, penelitian ini akan mengumpulkan data pertama, membersihkan data, dan kemudian mengubah data. Pembersihan data adalah proses membersihkan data, sedangkan Mengubah ukuran data untuk mencapai distribusi yang diharapkan dikenal sebagai transformasi data. Proses klasifikasi *Naive Bayes* akan terdiri dari dua fase, fase pelatihan dan fase pengujian. Karena *Naive Bayes* adalah metode pembelajaran yang diawasi, tahap pelatihan akan melibatkan pemahaman *Naive Bayes* tentang data untuk dapat membentuk model probabilitas. Selanjutnya, untuk memproses dan melakukan pengujian model, model proses data mining standar *Cross-Industry (CRISP-DM)* akan digunakan. Komponen CRISP-DM meliputi pemahaman bisnis, pemahaman data, persiapan data, modeling, evaluasi, dan penerapan. Hasil penelitian akan menunjukkan apa yang terjadi dan seberapa akurat penelitian itu.

### 2.2.1 CRISP DM

Pada langkah ini, metode CRISP-DM digunakan. Ini adalah metode yang banyak digunakan oleh para ahli untuk menggunakan model proses pengembangan data dan memecahkan masalah. Penelitian ini mencakup enam fase CRISP-DM, masing-masing digambarkan sebagai berikut:

#### 1. *Business Understanding (Pemahaman Bisnis)*

Tahap Pemahaman Bisnis, juga dikenal sebagai "Pemahaman Bisnis", berpusat pada tujuan dan kebutuhan, apakah itu terkait dengan bisnis, penelitian, atau kebutuhan lain. Selain

itu, tahapan ini menunjukkan pemahaman tentang masalah yang akan diteliti. Tujuan penelitian ini adalah untuk menggunakan algoritma *Naive Bayes* untuk mengkategorikan respons masyarakat terhadap Program Kartu Prakerja di *Twitter*. Selain itu, tingkat akurasi algoritma ini diukur untuk mengkategorikan sentimen negatif dan positif. [11].

## 2. *Data Understanding (Pemahaman Data)*

Pada tahap ini, pemahaman data juga dikenal sebagai pemahaman data membantu memberikan dasar analitik untuk penelitian dengan membuat ringkasan data dan menemukan masalah potensial. Data untuk Program Kartu Prakerja dikumpulkan dari 1.371 catatan *tweet* dari tahun 2023 melalui *Crawling Data Google Colab*. Catatan *tweet* memiliki enam atribut: *Username, Hastags, Time Stamp, Replies, Tweet Text, dan Retweet*. Data ini akan dibersihkan dan diubah sebelum diproses.

## 3. *Data Preparation*

Pada tahap selanjutnya, banyak tugas dilakukan, seperti persiapan data, yang mencakup semua tindakan yang diperlukan untuk membuat data akhir, atau data yang akan digunakan dalam alat pemodelan, dari data mentah awal. Pada tahap ini juga dilakukan pembersihan dan transformasi data [12].

## 4. *Permodelan (Modelling)*

Tahap berikutnya adalah melakukan pemodelan, yang merupakan hasil dari pelatihan dan pengujian data yang akan dilakukan. Alat pemodelan yang digunakan adalah *Rapidminer*. Pemodelan yang dilakukan adalah Algoritma *Naive Bayes* kemudian akan pakai sebagai pengukur tingkat akurasi dengan menggunakan *Confusion Matrix* [13].

## 5. *Evaluasi*

Pada tahap ini evaluasi yang bertujuan untuk menentukan suatu nilai dari model yang telah dibuat sebelumnya.

## 6. *Penyebaran (Deployment)*

Setelah pemodelan dan evaluasi proses data mining, tahap selanjutnya adalah penyebaran hasil penelitian [14].

### 2.2.2 *Metode Algoritma Naive Bayes*

Metode algoritma *Naive Bayes* terdiri dari serangkaian algoritma yang disusun berdasarkan *Teorema Bayes* seperti pada formula (1). Algoritma ini menggunakan prinsip probabilitas bersyarat dan asumsi independensi antar variabel penjelas. *Naive Bayes* digunakan dalam berbagai situasi, seperti klasifikasi dokumen, penyaringan pesan *spam*, dan sistem rekomendasi. Tipe algoritma *Naive Bayes* yang populer meliputi *Naive Bayes Gaussian, Naive Bayes Multinomial, dan Bernoulli*. Kelebihan algoritma ini adalah kecepatan aplikasi yang tinggi dan tingkat nilai error yang lebih rendah untuk dataset yang besar [15]

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (1)$$

Keterangan:

X adalah sampel data dengan label kelas (label) yang tidak diketahui

C; hipotesis bahwa X adalah data kelas (label);

P(X) adalah peluang dari sampel data yang diamati (probabilitas C); dan

P(X|C) adalah peluang yang didasarkan pada kondisi hipotesis.

Proses *Naive Bayes* digambarkan sebagai berikut:

1. Menemukan nilai peluang kasus baru dari setiap hipotesa yang memiliki kelas (label) yang ada di  $P(C_i)$ ,
2. Menemukan nilai akumulasi peluang dari setiap kelas  $P(X|C_i)$ ,
3. Menemukan nilai  $P(X|C_i) \times P(C_i)$ , dan
4. Menentukan kelas kasus baru.

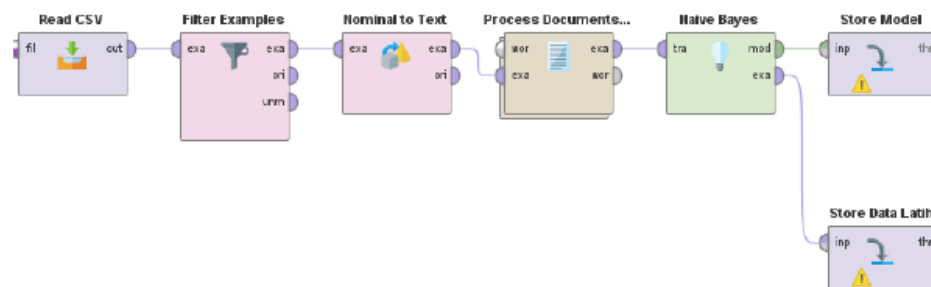
### 3. HASIL DAN PEMBAHASAN

#### 3.1 Metode Pengumpulan Informasi

Pertama, data dari *platform* media sosial *Twitter* dikumpulkan menggunakan alat pengumpul data *auth\_token*. *Dataset* publik yang akan digunakan untuk penelitian ini berasal dari 1.371 *tweet* pengguna *Twitter* yang dikumpulkan menggunakan *Google Colab Python* dengan kata kunci "Program Kartu Prakerja", dan kemudian difilter. Data yang dikumpulkan mencakup tanggapan pengguna terhadap *tweet* yang ditujukan kepada masyarakat tentang Program Kartu Prakerja.

#### 3.2 Metode Pelabelan Set Data

Penelitian ini menggunakan *Rapidminer* untuk secara otomatis melebeli data model atau kosong, dan 200 data akan dilabeli secara manual untuk digunakan sebagai data latih, dan 1008 data akan digunakan sebagai data model yang belum terlabeli. Gambar 2 dan 3 menunjukkan bahwa data diberi label "positif" dan "negatif".



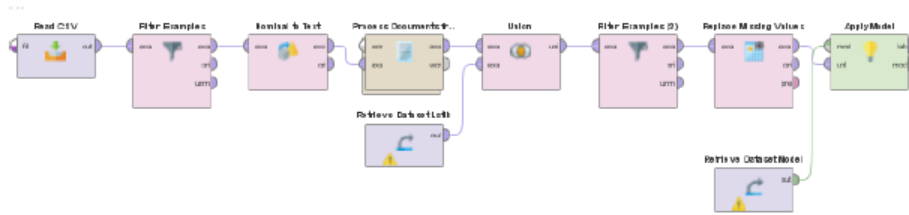
Gambar 2 Menghasilkan Data Latih dan Model



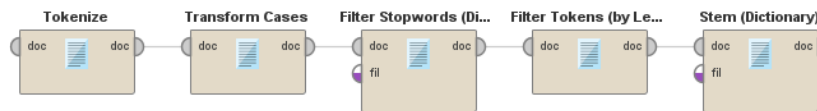
Gambar 3 Data Model dan Data Latih

Tujuan dari tahapan ini adalah untuk membangun sistem yang menggunakan informasi seperti label atau sentimen. Akibatnya, seperti yang ditunjukkan pada Gambar 4, Gambar 5, dan Gambar 6, operator union secara otomatis membeli dataset peneliti yang digunakan pada *Rapidminer Studio* setelah mengikuti beberapa kriteria:

1. *Tweets* yang menggunakan kata-kata positif seperti "baik", "kompeten", "menakjubkan", dan sebagainya akan diterima dengan baik.
2. *Tweet* yang mengatakan bahwa pengguna setuju bahwa honorer yang mengikuti prosedur Program Kartu Prakerja akan dianggap positif.



Gambar 4 Mempertimbangkan Label Data Kosong



Gambar 5 Proses Document From Data (Processing)

Result History

ExampleSet (Apply Model)

Open in Turbo Prep Auto Model

Filter (810 / 810 examples): all

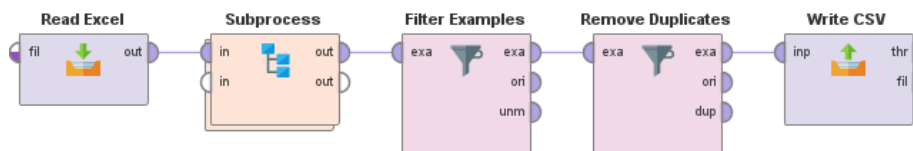
Row No.	Status	prediction(S...	confidence...	confidence...	text	Minggu	aamin	abis	acade
1	?	Positif	1	0	semangaatt...	0	0	0	0
2	?	Positif	1	0	tingkatkan ski...	0	0	0	0
3	?	Positif	1	0	kartu prakerja...	0	0	0	0
4	?	Positif	1	0	sertifikat das...	0	0	0	0
5	?	Negatif	0	1	vermuk siste...	0	0	0	0
6	?	Positif	1	0	pastinya gant...	0	0	0	0
7	?	Positif	1	0	pastinya gant...	0	0	0	0
8	?	Positif	1	0	prakerja bes...	0	0	0	0
9	?	Negatif	0	1	daya beli ma...	0	0	0	0
10	?	Positif	1	0	gampang dat...	0	0	0	0
11	?	Negatif	0	1	versi ganteng...	0	0	0	0
12	?	Positif	1	0	nelp prakerja ...	0	0	0	0
13	?	Negatif	0	1	prakerja dap...	0	0	0	0
14	?	Negatif	0	1	program kart...	0	0	0	0

ExampleSet (810 examples, 5 special attributes, 2,610 regular attributes)

Gambar 6 Prediksi Label Data

### 3.3 Tahap Cleansing Data

Selama proses sentimen analisis, penghapusan kata-kata yang tidak diperlukan dalam *tweet*, seperti *hashtag*, pernyataan, URL, dan karakter "@", adalah bagian dari proses pembersihan data. Subproses mengandung kumpulan data perubahan, termasuk perubahan URL, pernyataan, dan simbol. Salah satu contoh pengaturan adalah memfilter teks yang sudah ada atau terisi, menghilangkan duplikat, dan menghilangkan data yang sama seperti yang ditunjukkan pada Gambar 7 dan Gambar 8.

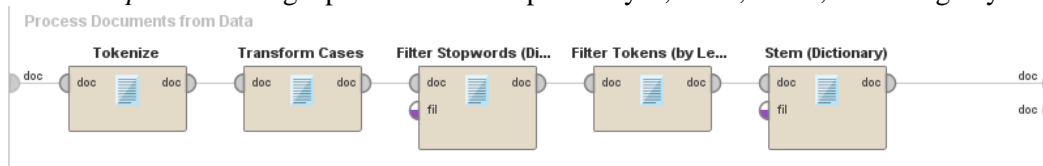


Gambar 7 Cleansing Data

1. *Tokenizing*: Membagi teks menjadi kata-kata, seperti "sedang mengerjakan skripsi", yang terdiri dari empat kata: "sedang", "mengerjakan", dan "skripsi".
2. *Folding case*: Proses mengubah data *tweet* menjadi *case lower* (huruf kecil).
3. *Stemming*: Membersihkan kata-kata imbuhan awalan dan akhiran yang ada dalam teks, seperti mengubah "mengerjakan" menjadi "kerja".



4. *Filter Tokens (by Length)*: Menghapus kata-kata yang memiliki kurang dari dua huruf dan lebih dari dua puluh lima huruf.
5. *Filter Stopwords*: Menghapus kata bantu seperti "saya", "dia", "aku", dan sebagainya.

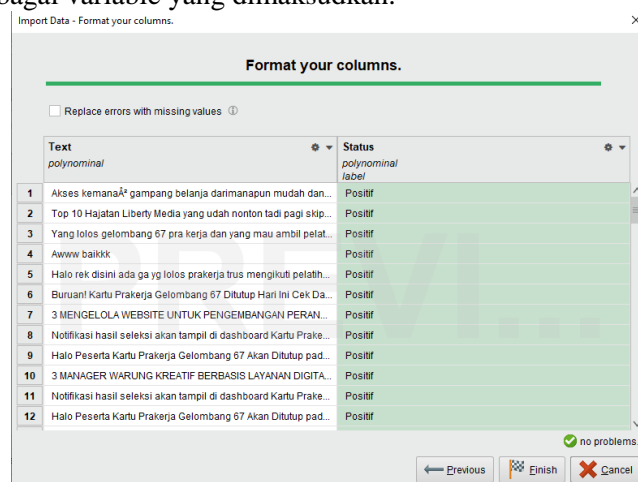


Gambar 8 *Process Document From Data (Processing)*

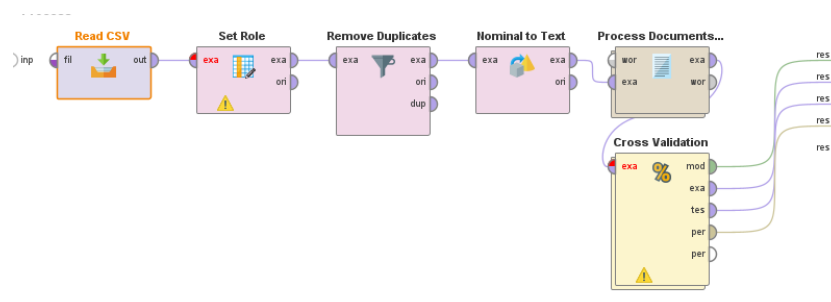
### 3.4 Hasil Akhir

Dalam proses *import* data ke aplikasi *RapidMiner* akan memasukkan *dataset* atau data *Excel* yang telah disiapkan, dan kemudian akan memasukkan label dan tipe data yang dipilih. Pada tahap pertama penelitian ini, setiap data harus diberi label dan diklasifikasikan menurut jenis datanya, seperti yang ditunjukkan dalam Gambar 9 dan Gambar 10, yang menunjukkan penerapan model. Selanjutnya, tipe label dan atributnya harus ditetapkan sesuai dengan atributnya, yaitu sebagai berikut:

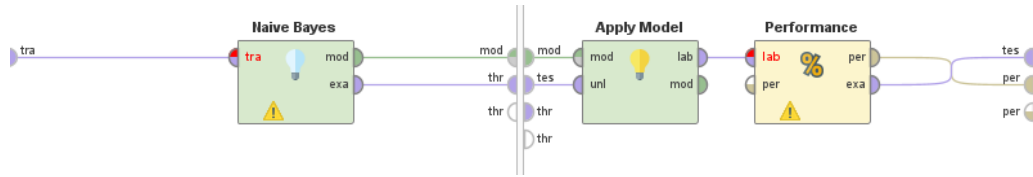
1. *Polynomial* adalah jenis data seperti *varchar* atau teks yang terdiri dari angka dan huruf.
2. *Binominal* adalah tipe data dua kategori, seperti Y/T, L/P, dan besar/kecil, dll.
3. *Atribut berfungsi sebagai prediktor atau variable prediksi*.
4. *label berfungsi sebagai variable yang dimaksudkan*.



Gambar 9 Penentuan Label Dan Tipe Data



Gambar 10 *Model Algoritma Naïve Bayes*



Gambar 11 Menghubungkan *Port* Pada Tiap Operator Pada *Naïve Bayes*

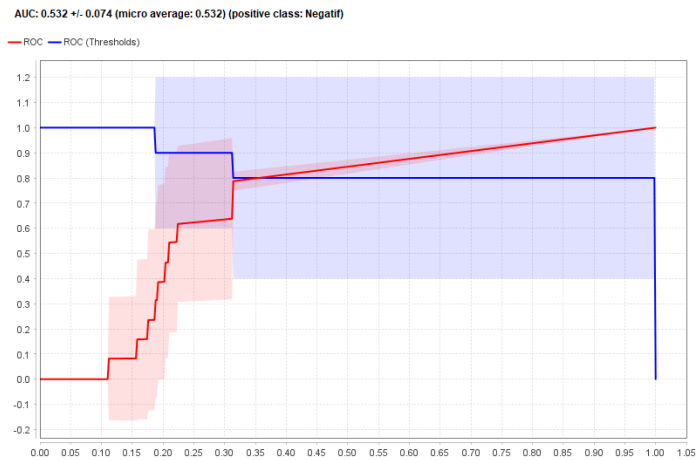
Hasil perhitungan akurasi data latih dengan metode *Naïve Bayes* menunjukkan nilai akurasi 77,58%. Dari 1.010 data uji, 423 Seperti yang ditunjukkan pada Gambar 11, data menunjukkan sentimen negatif dan 596 sentimen positif.

accuracy: 77.58% +/- 4.81% (micro average: 77.58%)

	true Positif	true Negatif	class precision
pred. Positif	502	94	84.23%
pred. Negatif	132	280	67.96%
class recall	79.18%	74.87%	

Gambar 12 *Confusion Matix* Algoritma *Naïve Bayes*

Hasil uji coba menunjukkan akurasi *Naïve Bayes* sebesar 77,58%, seperti yang ditunjukkan pada Gambar 12. Selain itu, *Vector Performance* menunjukkan jumlah dan jenis data mana yang sesuai atau tidak sesuai antara hasil uji coba dan perhitungan *Naïve Bayes*. Hasil perhitungan dikategorikan menjadi benar positif, benar negatif, benar negatif, dan benar negatif.

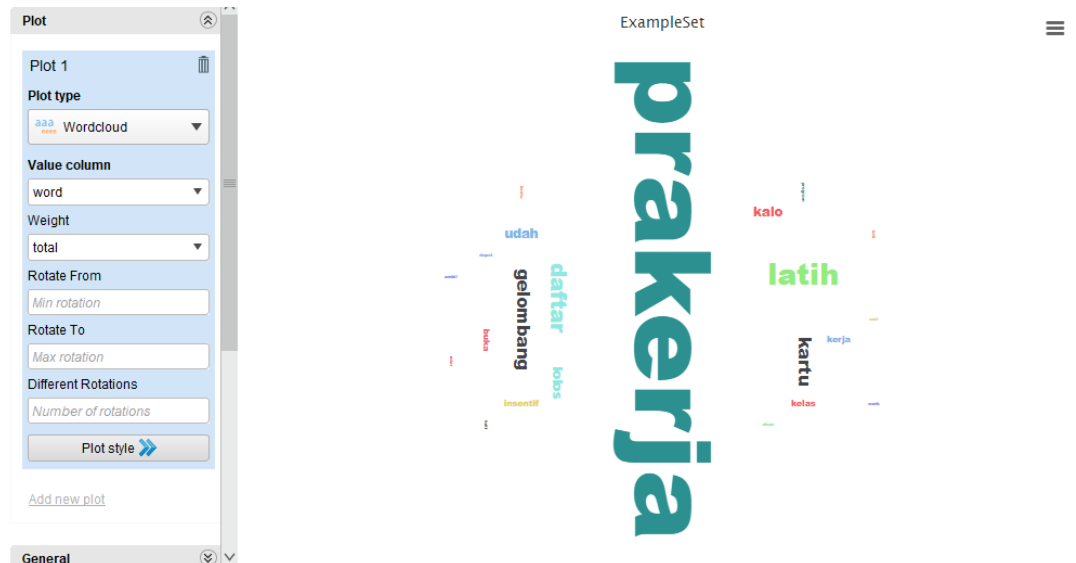


Gambar 13 Hasil Nilai *Curva ROC*

Seperti yang ditunjukkan pada Gambar 13, Berdasarkan hasil, *Algoritma Naïve Bayes* dianggap memiliki klasifikasi yang baik dengan *kurva ROC* sebesar 0.532.

### 3.2 Hasil *World Cloud*

Dari 1.008 data pengujian, 374 menunjukkan status negatif dan 634 menunjukkan status positif, hasil akhir *Naïve Bayes* menunjukkan nilai akurasi sebesar 77,58% Seperti yang ditunjukkan pada Gambar 14, Hasil ini diimplementasikan dengan teknik *Naïve Bayes* dan didasarkan pada data *Twitter*.



Gambar 14 World Cloud

#### 4. KESIMPULAN

Sebagai kesimpulan dari hasil penelitian dilakukan untuk menemukan hasil Matriks Konflik untuk Analisis Sentimen Naïve Bayes, dapat dikatakan bahwa:

- Dengan *RapidMiner Studio*, metode *Naive Bayes* dapat digunakan untuk melakukan sentimen analisis dengan menggunakan proses *Crawling*, *Labeling*, *Cleaning*, *Preprocessing*, dan *Klasifikasi* pada data *Twitter*.
- Hasil analisis persepsi menunjukkan bahwa opini masyarakat tentang Program Kartu Prakerja 634 Dari 1.008 data uji, 374 menunjukkan hasil positif. Ini mengindikasikan bahwa metode *Naive Bayes* memiliki nilai akurasi 77,58%.

#### 5. SARAN

Hasil dari analisis yang dilakukan selama penelitian memberikan saran untuk pengembangan dan penelitian lebih lanjut, seperti:

- Untuk meningkatkan proses prediksi sentimen dan mendapatkan hasil yang lebih baik dan akurat, diharapkan untuk meningkatkan jumlah data uji.
- Teknik data mining yang tidak sama dengan yang sebelumnya digunakan dapat digunakan untuk melakukan penelitian serupa.

#### DAFTAR PUSTAKA

- [1] S. Syahrani dan A. Gunawan, "Global: Jurnal Lentera BITEP Efektivitas Program Kartu Prakerja dalam Membangun Sumber Daya Manusia dan Pengembangan Skill," 2023.
- [2] B. Zaki dan T. Kartika Pertiwi, "Pengaruh Program Kartu Prakerja terhadap Penyerapan Tenaga Kerja Indonesia pada Masa Pandemi Covid - 19," Apr 2023.
- [3] E. M. F. C. M. T. Gilbert Luis Ondang, "Efektifitas Program Kartu Prakerja bagi fresh graduate di kota manado," 2022.
- [4] N. Hendrastuty, A. Rahman Isnain, dan A. Yanti Rahmadhani, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," 2021.

- 
- [5] W. P. Anggraini dan M. S. Utami, “KLASIFIKASI SENTIMEN MASYARAKAT TERHADAP KEBIJAKAN KARTU PEKERJA DI INDONESIA,” *Faktor Exacta*, vol. 13, no. 4, hlm. 255, Feb 2021.
- [6] S. Syahrani dan A. Gunawan, “Global : Jurnal Lentera BITEP Efektivitas Program Kartu Prakerja dalam Membangun Sumber Daya Manusia dan Pengembangan Skill,” 2023.
- [7] F. B. Dinda Indah Nabila, “Pengaruh Motivasi, Komitmen Dan Lingkungan Belajar Terhadap Pelatihan Program Kartu Prakerja,” 2023.
- [8] D. Anggraeni, R. A. Ghofur, dan S. Hilal, “Efektivitas Program Kartu Prakerja dalam Membangun Sumber Daya Manusia Pasca Pandemi Covid-19,” *Jurnal Informatika Ekonomi Bisnis*, hlm. 890–893, Agu 2023.
- [9] R. A. Ellya, H. Sekolah, T. Ekonomi, dan D. B. Syariah, “Analisis Kebijakan Program Kartu Prakerja Sesuai Dengan Perspektif Ekonomi Islam di Kelurahan Satu Ilir Palembang.”
- [10] R. I. Wijayanti, A. Humardhiana, I. Syekh, dan N. Cirebon, “Optimalisasi Program Kartu Prakerja Dengan Pelatihan Branding Strategy,” vol. 2, no. 1, hlm. 109, 2020.
- [11] D. Rahman dan K. Rahman, “Jurnal Kajian Pemerintah (JKP) Journal of Government, Social and Politics EFEKTIVITAS KEBIJAKAN PEMERINTAH TENTANG PENGEMBANGAN KOMPETENSI KERJA MELALUI PROGRAM KARTU PRAKERJA STUDI PROGRAM KARTU PRAKERJA DALAM MENGATASI PENGANGGURAN DI KABUPATEN TEBO PROVINSI JAMBI 1,” 2021.
- [12] F. S. Kwure dan A. Susiantoro, “EFEKTIVITAS BANTUAN KARTU PRAKERJA TERHADAP PEREKONOMIAN MASYARAKAT DI MASA COVID-19 DI KOTA SURABAYA,” *Jurnal Penelitian Administrasi Publik*, 2023.
- [13] S. D. Yana, “Efektifitas Program Kartu Prakerja Dalam Membangun Sumberdaya Manusia di Tengah Pandemi,” 2021.
- [14] Tasmilah, “Pengaruh Kartu Prakerja dalam Menciptakan Wirausaha Baru pada Generasi Y dan Z,” *Jurnal Ekonomi Indonesia* •, vol. 12, hlm. 89–107, 2023.
- [15] E. Andina, “Peran Pemerintah Daerah dalam Implementasi Program Kartu Prakerja di Provinsi Jawa Barat,” *Aspirasi: Jurnal Masalah-masalah Sosial*, vol. 13, no. 1, Jun 2022.