

Analisis Sentimen Penilaian Masyarakat Indonesia terhadap *GeNose* pada Komentar *Youtube* Menggunakan Metode *Naïve Bayes*

Michael Jonathan^{1*}, Yessica Nataliani^{1*}

¹Jurusan Sistem Informasi–Fakultas Teknologi Informasi–Universitas Kristern Satya Wacana, Indonesia

Email: 682018137@student.uksw.edu*, yessica.nataliani@uksw.edu*

ABSTRAK

Saat ini dunia sedang menghadapi virus yang benar-benar mematikan yaitu COVID-19 karena virus ini menyebar dengan sangat cepat dan menyerang paru-paru. Banyak orang yang meninggal dunia dikarenakan sulit untuk bernafas. Khususnya di Indonesia, banyak masyarakat Indonesia yang terkena COVID-19 dan banyak yang sampai meninggal dunia. Banyak daerah yang ditetapkan sebagai zona merah yang menandakan bahwa daerah tersebut beresiko tinggi. Oleh karena itu universitas berbondong-bondong untuk menciptakan alat yang dapat meminimalisir penyebaran COVID-19. Salah satunya adalah Universitas Gadjah Mada (UGM) yang menciptakan alat yang dapat mendeteksi COVID-19 yaitu *GeNose*. Masyarakat memiliki beberapa argumen pro dan kontra dengan adanya *GeNose*. Untuk itu diperlukan penelitian untuk mengetahui apakah *GeNose* dapat berguna terhadap masyarakat Indonesia dalam meminimalisir virus tersebut atau tidak. Penelitian dilakukan menggunakan metode *Naïve Bayes*. Hasil klasifikasi menghasilkan *accuracy* mencapai 90.70% dengan kategori komentar bersifat negatif. Nilai *recall* sebesar 87.60% dan *precision* sebesar 93.39%. Data tersebut membuktikan bahwa *GeNose* kurang maksimal dalam meminimalisir COVID-19.

INFO ARTIKEL

Diterima
Diterima setelah revisi
Tersedia online

Kata Kunci:

COVID-19
GeNose
Naïve Bayes

ABSTRACT

Currently the world is facing a truly deadly virus, namely COVID-19 the virus spreads very quickly and attacks the lungs. So many people die because it is difficult to breathe. Especially in Indonesia, many Indonesians have been affected by COVID-19 and many have died. So many areas are designated as red zones which indicate that the area is at high risk. Therefore, many flocked in droves to be able to create tools that can minimize the spread of COVID-19. One of them is Universitas Gadjah Mada (UGM) which invented a tool that can detect COVID-19, namely *GeNose*. However, the public has several pros and cons arguments for *GeNose*. This research wants to find out whether *GeNose* can be useful for the Indonesian people in minimizing the virus in Indonesia or not, by using the *Naïve Bayes* method. The classification results give an accuracy of 90.70% with the comments category of negative. The recall value is 87.60% and the precision is 93.39%. These data prove that *GeNose* is not optimal in minimizing COVID-19.

ARTICLE INFO

Accepted
Accepted after revision
Available online

Keywords:

COVID-19
GeNose
Naïve Bayes

1. PENDAHULUAN

Kasus COVID-19 pertama kali menyebar di Wuhan, Provinsi Hubei, Cina pada Desember 2019. COVID-19 dapat berkembang dengan waktu yang singkat dan telah menyebar ke provinsi lain yang terdapat di Cina, bahkan dengan waktu yang singkat yaitu kurang dari satu bulan COVID-19 sudah menyebar kenegara lain. Oleh karena itu pandemi COVID-19 sekarang ini menjadi perhatian seluruh negara salah satunya adalah negara Indonesia [1]. Warga negara Indonesia yang terkena virus COVID-19 masih terus bertambah. Berdasarkan berita yang diumumkan oleh Kemenkes RI, pada bulan Juli 2021 tercatat 2.615.529 kasus penyebaran yang terkonfirmasi. Hal ini menyebabkan banyak universitas berbondong-bondong untuk dapat menciptakan alat yang berguna untuk meminimalisir penyebaran COVID-19. salah satu

universitas yang berhasil menciptakan alat yang dapat mendeteksi COVID-19 yaitu Universitas Gadjah Mada (UGM) yang menciptakan alat yang dapat mendeteksi COVID-19 melalui nafas seseorang yang diberi nama *GeNose*. Alat ini bekerja dengan mendeteksi nafas seseorang dan hasilnya akan keluar dalam beberapa menit [2]. Masyarakat Indonesia memiliki penilaian tersendiri terhadap *GeNose* yang biasanya di-posting di media sosial, salah satunya *Youtube*.

Untuk mengetahui penilaian masyarakat Indonesia tentang *GeNose*, maka dilakukan analisis sentimen. Analisis sentiment memiliki tujuan untuk melakukan analisis terhadap pendapat yang merupakan pengekspresian opini-opini orang tentang suatu topik yang terdapat di media sosial. Metode ini sering dilakukan untuk mendapatkan kesimpulan berdasarkan

pemikiran orang terhadap suatu produk, preferensi politik, pendapat tentang film yang telah dirilis serta prediksi pendapatan film, dan lainnya [3].

Berdasarkan latar belakang permasalahan tersebut, oleh karena itu penelitian tentang analisis sentimen penilaian masyarakat terhadap *GeNose* akan dilakukan dengan menggunakan metode *Naïve Bayes*. Metode ini mempunyai kelebihan dari metode yang lain yaitu perhitungan yang cepat, sederhana, dan memiliki akurasi yang cukup tinggi [4]. Dalam menggunakan metode ini tidak memerlukan data yang banyak untuk menghasilkan estimasi parameter dalam proses penelitian. Metode ini mudah digunakan karena mempunyai alur perhitungan yang mudah dimengerti dan tidak penjang. Penelitian memiliki tujuan yaitu mengetahui pendapat masyarakat tentang *GeNose* apakah *GeNose* dapat membantu masyarakat dalam meminimalisir penyebaran *COVID-19* atau bahkan membuat resah masyarakat. Penelitian ini bersumber berdasarkan pendapat masyarakat yang terdapat pada komentar yang ada di *Youtube*, khususnya *channel Kompas TV* dan *Official iNews*. Karena setiap komentar masyarakat memiliki bobot penilaian masing-masing, oleh karena itu penelitian dilakukan dengan mengelompokan komentar masyarakat Indonesia tentang *GeNose* menjadi dua kelompok yaitu positif dan negatif. Setelah itu akan dilakukan pengklasifikasian menggunakan metode *Naïve Bayes*.

Penelitian terdahulu merupakan penelitian yang sudah diteliti sebelumnya yang berguna untuk mendapatkan bahan untuk perbandingan dan acuan dalam pembuatan penelitian. Untuk menghindari adanya anggapan kesamaan penelitian terdahulu dengan penelitian yang sedang diteliti. Beberapa penelitian terdahulu tentang pengklasifikasian *Naïve Bayes* dijelaskan sebagai berikut.

Penelitian tentang pengklasifikasian pelanggan yang dilakukan menggunakan metode *Naïve Bayes*. Hasil dari penelitian berupa tabel *confusion matrix*, untuk menghitung nilai *precision*, *recall*, *accuracy*. Nilai *precision* yang didapat adalah sebesar 100%, *recall* sebesar 91%, dan *accuracy* sebesar 92% [5].

Penelitian tentang prediksi kelulusan dengan menggunakan metode *Naïve Bayes*. Hasil dari penelitian adalah berupa kelompok lulus pada masing-masing program studi dan didapatkan hasil bahwa kelompok lulus pada pilihan pertama lebih tinggi dibandingkan kelompok lulus pada pilihan kedua serta pilihan tidak lulus [6].

Penelitian tentang masyarakat miskin dengan metode pengklasifikasian dan algoritma *Naïve Bayes*. Hasil dari penelitian berupa *form* yang menunjukkan hasil akhir dari sistem, yang berupa hasil klasifikasi setiap objek menggunakan metode yang diusulkan. Selain itu terdapat proses pencetakan dari hasil klasifikasi sebagai laporan kepada pimpinan [4].

Jurnal tentang penerapan metode *Naïve Bayes* terhadap bantuan sosial keluarga prasejahtera dengan menggunakan metode klasifikasi dan *Naïve Bayes*. Hasil penelitian adalah data keluarga penerima bantuan sosial

pra-sejahtera yang dapat diperoleh dengan mudah. Hasil tersebut didapat dengan tidak diperlukannya pencatatan ulang dari data sebelumnya, tetapi dengan menggabungkan data yang sudah ada. Dengan penelitian ini maka keluarga yang layak mendapatkan bantuan dan yang tidak dapat terlihat, sehingga pemberian bantuan sosial keluarga pra-sejahtera dapat terlaksana dengan baik [7].

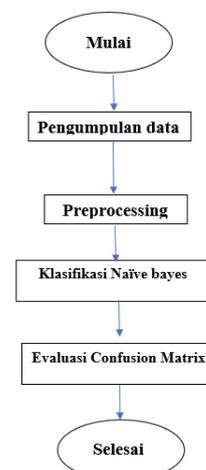
Jurnal tentang perbandingan algoritma untuk klasifikasi analisis sentimen terhadap *GeNose* pada media sosial *Twitter* menggunakan metode *data mining*. Hasil penelitian adalah algoritma *Naïve Bayes* memiliki performa yang cukup baik dengan *accuracy* 72,36%, sedangkan algoritma *k-Nearest Neighbor* menghasilkan *accuracy* 64,22%, dan algoritma *Decision Tree* dengan *accuracy* 65,15% [2].

Penelitian tentang klasifikasi jurnal ilmu komputer berdasarkan pembagian pada *web science* menggunakan metode *text mining* dengan bantuan RapidMiner. Hasil dari penelitian didapat bahwa klasifikasi dengan *Naïve Bayes Classifier* memberikan nilai performa yang lebih tinggi dibandingkan metode *Support Vector Machine*. Ukuran performansi yang digunakan adalah *recall*, *precision*, *F-measure*, dan *accuracy*, dimana metode *Naïve Bayes* mendapatkan nilai terbaik. Berdasarkan nilai-nilai tersebut, dapat dilakukan otomatisasi pengklasifikasian jurnal-jurnal lainnya menggunakan metode *Naïve Bayes* [8].

Perbedaan dengan penelitian terdahulu di atas yaitu pada objek yang dibahas dan periode pengamatan. Selain itu penelitian ini juga memiliki keunggulan salah satunya adalah kasus yang diambil adalah kasus yang terbaru saat ini di negara Indonesia. Jadi penelitian ini juga dapat membantu negara Indonesia dalam menyelesaikan kasus permasalahan. Selain itu penelitian ini mengambil data melalui komentar pada *Youtube*.

2. METODE PENELITIAN

Beberapa tahapan Penelitian yang dibuat secara sederhana dengan menggunakan diagram yang dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

Berikut penjelasan berdasarkan alur penelitian pada Gambar 1. Tahapan pertama terdapat tahapan pengumpulan data, dimana pada tahapan ini akan dilakukan proses pengambilan data dari *Youtube*. Tahap selanjutnya adalah *preprocessing*, dimana pada tahapan ini dilakukan pemfilteran data secara manual. Selanjutnya adalah tahap klasifikasi *Naive Bayes*, tahapan ini adalah proses pengklasifikasian data dengan bantuan aplikasi *rapidminer*. Lalu terdapat tahap evaluasi *confusion matrix*, pada tahap ini akan dilakukan evaluasi data dengan menggunakan rumus *confusion matrix*.

Pengumpulan Data

Pengumpulan data merupakan tahap untuk mengumpulkan data [3]. Data yang diambil untuk penelitian berasal dari komentar sosial media *Youtube* sesuai topik permasalahan yang diambil. Data yang digunakan sebagai penelitian berasal dari dua *channel Youtube* yaitu Kompas TV dan Official iNews.

Pada tahap ini dilakukan dengan menggunakan metode *web scraping* pada *channel Youtube* yang dipilih. Pada *web Scraping* menggunakan bahasa pemrograman yaitu Javascript dan dengan bantuan *google spreadsheet*.

Preprocessing

Tahap *preprocessing* data merupakan proses mempersiapkan data untuk menjadi pemodelan. *Preprocessing* memiliki tahapan proses pembentukan data dan pembersihan data.

Tahap pertama pada *preprocessing* adalah *cleansing* data dengan cara manual. Data yang tidak termasuk ke dalam topik permasalahan yang diambil akan dihapus. Kegunaan dari tahap ini adalah membersihkan data dari data yang tidak sesuai dengan topik permasalahan. Selain itu dapat mempermudah proses pengklasifikasian [9].

Selanjutnya akan dilakukan beberapa tahapan yaitu: menghapus tanda baca serta *hashtag*, menjadikan huruf kecil pada kalimat, menghilangkan *stopword*, menghilangkan kata imbuhan, dan mengubah kata menjadi kata baku [10].

Klasifikasi *Naive Bayes*

Sebelum melakukan tahap pengklasifikasian data akan dilakukan pengelompokan data dengan memberikan label pada tiap data. Pelabelan data terdapat 2 jenis label yaitu positif dan negatif. Dengan melakukan tahap ini dapat membantu menentukan kelas dari objek yang belum diketahui labelnya [11].

Selanjutnya adalah tahapan pengklasifikasi menggunakan klasifikasi *Naive Bayes*. Metode ini merupakan suatu metode yang sangat berguna untuk penghitungan klasifikasi karena *Naive Bayes* merupakan pengklasifikasian yang terdapat metode probabilitas dan statistik [12]. Metode ini dikemukakan oleh ilmuwan Inggris Thomas Bayes, yang diciptakan

untuk melihat peluang di masa yang akan datang didasarkan pada pengalaman sehingga menjadi sangat berguna untuk melakukan prediksi yang akan datang [13]. Berikut rumus persamaan yang digunakan dalam menggunakan metode *Naive Bayes* dapat dilihat pada Rumus (1).

$$P(c|X) = \frac{P(X|c).P(c)}{P(X)} \quad (1)$$

dimana

- $P(c|X)$: Probabilitas hipotesis menurut kondisi
- $P(c)$: Probabilitas hipotesis
- $P(X|c)$: Probabilitas yang didasari oleh kondisi pada hipotesis
- $P(X)$: Probabilitas c

Evaluasi dengan *Confusion Matrix*

Tahapan ini berguna untuk menganalisis hasil dari klasifikasi sentimen berdasarkan hasil aktual dan prediksi. Dengan adanya hasil perbedaan antara actual dan prediksi, oleh karena itu dapat diketahui hasil pengklasifikasian antara prediksi dan actual. Perhitungan dengan menggunakan *confusion matrix* dapat menampilkan prediksi positif dan negatif terhadap kelas negatif maupun kelas positif [15]. Tabel *confusion matrix*-nya dapat dilihat seperti pada Tabel 1.

Tabel 1. *Confusion Matrix*

	Positif	Negatif
Pred. Positif	TP	FP
Pred. Negatif	FN	TN

dimana:

- TP (*True Positive*) merupakan jumlah data kelas positif dan terklasifikasi positif.
- TN (*True Negative*) merupakan jumlah data kelas negatif dan terklasifikasi negatif.
- FN (*False Negative*) merupakan jumlah data kelas negatif, tetapi terklasifikasi salah.
- FP (*False Positive*) merupakan jumlah data kelas positif, tetapi terklasifikasi salah.

Melalui data tersebut, maka didapatkan data lain yang berguna untuk melihat performa pada suatu model, di antaranya [14]:

- $Accuracy$ = menggambarkan seberapa akurat model dalam mengklasifikasikan dengan benar. Formula $accuracy$ dapat dilihat pada Rumus (2).

$$Accuracy = \frac{TP+TN}{Total} \quad (2)$$

- $Precision$ = menggambarkan akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model. Formula $precision$ dapat dilihat pada Rumus 3.

$$Precision = \frac{TP}{FP+TP} \quad (3)$$

- $Recall$ (*Sensitivity/True Positive Rate*) = menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi. Formula $recall$ dapat dilihat pada Rumus (4).

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

3. HASIL DAN PEMBAHASAN

Pengumpulan Data

Pengumpulan data diambil berdasarkan komentar masyarakat pada media *Youtube* dengan topik “*GeNose*” yang diambil dari beberapa sumber yaitu Kompas TV yang di-*posting* pada tanggal 3 Februari 2021, 5 Februari 2021, dan 6 Februari 2021, serta *channel* Official *iNews* yang diposting pada tanggal 25 Juni 2021. Pengumpulan data pada komentar *Youtube* berdasarkan komentar negatif dan positif sesuai dengan parameter yang disebutkan pada metode penelitian. Hasil dari proses pengambilan data dalam format Excel.

Data komentar *Youtube* yang hanya diambil sebanyak 1000 data. Tahap pengumpulan data dilakukan dengan bantuan *software* seperti Google Document dan pemrograman Javascript. Pemrograman ini digunakan untuk mengambil data dengan mudah dalam cakupan yang banyak pada komentar *Youtube*. Berikut *coding* yang digunakan pada Javascript untuk mengambil data komentar *Youtube*.

```
function scrapeCommentsWithoutReplies() {
    var ss =
    SpreadsheetApp.getActiveSpreadsheet();
    var
    result=[['Name', 'Comment', 'Time', 'Likes', 'Reply
    Count']];
    var vid =
    ss.getSheets()[0].getRange(1,1).getValue();
    var nextPageToken=undefined;

    while(1){
    var data =
    YouTube.CommentThreads.list('snippet',
    {videoId: vid, maxResults: 100, pageToken:
    nextPageToken})

    nextPageToken=data.nextPageToken
    //console.log(nextPageToken);
    for (var row=0; row<data.items.length; row++)
    {
    result.push([data.items[row].snippet.topLevelC
    omment.snippet.authorDisplayName,

    data.items[row].snippet.topLevelComment.snippe
    t.textDisplay,

    data.items[row].snippet.topLevelComment.snippe
    t.publishedAt,

    data.items[row].snippet.topLevelComment.snippe
    t.likeCount,

    data.items[row].snippet.totalReplyCount]);
    }
    if(nextPageToken =="" ||
    typeof nextPageToken === "undefined"){
    break;
    }
    }

    var newSheet=ss.insertSheet(ss.getNumSheets())
    newSheet.getRange(1,
    1,result.length,5).setValues(result)
}
```

Preprocessing

Setelah tahap pengumpulan data, selanjutnya dilakukan proses *preprocessing* yaitu proses pra-pengolahan data. Langkah pertama yang akan dilakukan pada tahap *preprocessing* adalah *cleansing* data atau pembersihan data. Data yang diperoleh setelah melalui tahap *cleansing* sebanyak 484 data. Contoh data sebelum *cleansing* data yang ada pada Tabel 2 dan sesudah langkah *cleansing* yang ada pada dapat dilihat pada Tabel 3.

Tabel 2. Contoh Data Sebelum *Cleansing* Data

Name	Comment
macaolao xxx	Gk akurat ntr kenak ada virus yg bukan covid dikatakan covid.
Arfino Fino	Alahhhh itu cama buat bisniss smuanya buat bisniss....saya tanya kalo sudah di suntik vaksin ap hasilnya bisa positif ap negatip
Afiq Asyrofi	Genose banyak yg gak akurat Antigen positif tapi dg trik2 tertentu genose bisa jadi negatif
Madi nug	Subsidi PLTA ke pelanggan biar pada masuk tangerang selatan..
Shandy	Yang aku pikirkan genose muridnya saitama :)
Yoake	Genose anime one Punch Man 🤩

Tabel 3. Contoh Data Setelah *Cleansing* Data

Name	Comment
macaolao xxx	Gk akurat ntr kenak ada virus yg bukan covid dikatakan covid.
Arfino Fino	Alahhhh itu cama buat bisniss smuanya buat bisniss....saya tanya kalo sudah di suntik vaksin ap hasilnya bisa positif ap negatip
Afiq Asyrofi	Genose banyak yg gak akurat Antigen positif tapi dg trik2 tertentu genose bisa jadi negatif

Tahapan kedua yang dilakukan adalah menghilangkan tanda baca yaitu proses menghilangkan tanda baca, *hashtag*, dan *emoticon* yang tidak diperlukan. Contoh data sebelum melakukan *preprocessing* yang ada pada tabel 4 dan sesudah hasil *preprocessing* tahap 2 yang ada pada Tabel 5.

Tabel 4. Tabel sebelum melakukan *preprocessing* tahap 2

Name	Comment
------	---------

Analisis Sentimen Penilaian Masyarakat Indonesia terhadap GeNose pada Komentar Youtube Menggunakan Metode Naive Bayes

d'Cartesian: Jurnal Matematika dan Aplikasi, Vol. 11, No. 1, (Maret 2022): 1-8

macaolao xxx	Gak akurat nanti ada virus yg bukan covid dikatakan covid.
Faris Van java	Ribet banget dah mau pergi aja 😂😂
Arfino Fino	cuma buat bisnis smuanya buat bisnis....saya tanya kalo sudah di suntik vaksin ap hasilnya bisa positif apa negatif
Afiq Asyrofi	Genose banyak yang gak akurat Antigen positif tapi dengan trik2 tertentu genose bisa jadi negatif
Ginanjari Miftah	Apa cuma gue yang mikir disini. Jangan ada tes tes an covid gitu... Tutup berita covid, hidup kek dulu lagi..., tanpa masker... Jangan takut virus atau apalah... Pada dasarnya yang meninggal tiap hari juga banyak... Tapi apa boleh buat gue bukan presiden 😂😂😂😂 Sorry lagi Halu gabut
Ricky Rocks	iya sih masker engga dibuka. tapi tuh pipet dipegang ibunya abis itu dimasukin ke hidung reporternya. hadeh kocak wkwkwk

Tabel 5. Contoh Tabel Sudah melakukan *Preprocessing* Tahap 2

Name	Comment
macaolao xxx	Gak akurat nanti ada virus yg bukan covid dikatakan covid
Faris Van java	Ribet banget dah mau pergi aja
Arfino Fino	cuma buat bisnis smuanya buat bisnis saya tanya kalo sudah di suntik vaksin ap hasilnya bisa positif apa negatif
Afiq Asyrofi	Genose banyak yang gak akurat Antigen positif tapi dengan trik tertentu genose bisa jadi negatif
Ginanjari Miftah	Apa cuma gue yang mikir disini. Jangan ada tes tes an covid gitu Tutup berita covid hidup kek dulu lagi tanpa masker Jangan takut virus atau apalah Pada dasarnya yang meninggal tiap hari juga banyak Tapi apa boleh buat gue bukan presiden Sorry lagi Halu gabut
Ricky Rocks	iya sih masker engga dibuka tapi tuh pipet dipegang ibunya abis itu dimasukin ke hidung reporternya hadeh kocak wkwkwk

Langkah selanjutnya adalah proses menjadikan semua huruf menjadi huruf kecil. Contoh data sebelum melakukan *preprocessing* tahap 3 yang ada pada tabel 5 dan sesudah *preprocessing* tahap 3 yang ada pada Tabel 6.

Tabel 6. Contoh Tabel Sudah Melakukan *Preprocessing* Tahap 3

Name	Comment
macaolao xxx	gak akurat nanti ada virus yg bukan covid dikatakan covid
Faris Van java	ribet banget dah mau pergi aja
Arfino Fino	cuma buat bisnis smuanya buat bisnis saya tanya kalo sudah di suntik vaksin ap hasilnya bisa positif apa negatif
Afiq Asyrofi	genose banyak yang gak akurat antigen positif tapi dengan trik tertentu genose bisa jadi negatif
Ginanjari Miftah	apa cuma gue yang mikir disini jangan ada tes tes an covid gitu tutup berita covid hidup kek dulu lagi tanpa masker jangan takut virus atau apalah pada dasarnya yang meninggal tiap hari juga banyak tapi apa boleh buat gue bukan presiden sorry lagi halu gabut
Ricky Rocks	iya sih masker engga dibuka tapi tuh pipet dipegang ibunya abis itu dimasukin ke hidung reporternya hadeh kocak wkwkwk

Tahap selanjutnya adalah menghapus *stopword* yang merupakan tahapan menghilangkan kata-kata yang tidak penting pada kalimat komentar tersebut. *Stopword* berguna untuk menghilangkan kata dan, atau, mungkin, ini, dan itu. Contoh data sebelum melakukan *preprocessing* tahap ke 4 yang ada pada tabel 6 dan sesudah *preprocessing* tahap 4 yang ada pada Tabel 7

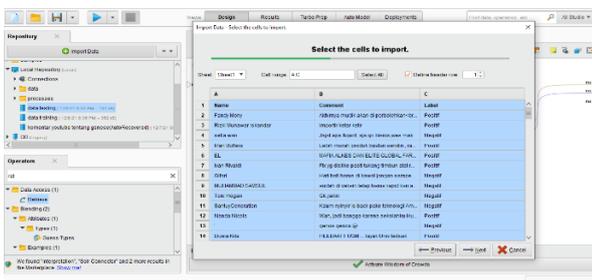
Tabel 7. Contoh Tabel Sudah Melakukan *Preprocessing* Tahap 4

Name	Comment
macaolao xxx	gak akurat nanti ada virus yg bukan covid dikatakan covid
Faris Van java	ribet banget dah mau pergi aja
Arfino Fino	cuma buat bisnis smuanya buat bisnis saya tanya kalo sudah di suntik vaksin hasilnya bisa positif apa negatif
Afiq Asyrofi	genose banyak yang gak akurat antigen positif tapi dengan trik tertentu genose bisa jadi negatif
Ginanjari Miftah	apa cuma gue yang mikir disini jangan ada tes tes an covid gitu tutup berita covid hidup kek dulu lagi tanpa masker jangan takut virus pada dasarnya yang meninggal tiap hari juga banyak tapi apa boleh buat gue bukan presiden sorry lagi halu gabut
Ricky Rocks	iya sih masker engga dibuka tapi tuh pipet dipegang ibunya abis itu dimasukin ke hidung reporternya hadeh kocak

Klasifikasi Naive Bayes

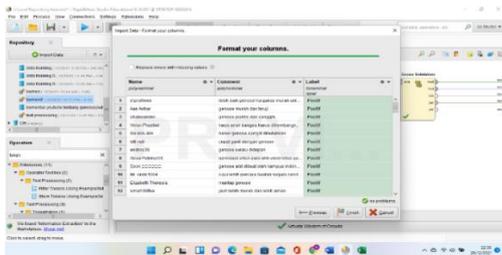
Sebelum pengklasifikasian terdapat tahap pelabelan data actual yang dilakukan secara manual yaitu dengan memberikan label positif jika komentar tersebut memiliki makna positif dan memberikan label negatif jika komentar tersebut memiliki makna negatif. Langkah selanjutnya setelah memberikan label adalah memberikan filter yang terdapat pada *Toolbar* pada Excel untuk membantu proses pengklasifikasian. Data yang diperoleh setelah melalui tahap pelabelan adalah 228 positif dan 256 negatif.

Tahap selanjutnya adalah *Naïve Bayes* dengan bantuan *RapidMiner*. Tahapan pada *RapidMiner* terdiri dari *read excel*, *smote upsampling*, dan *cross validation*. Yang harus dilakukan pertama kali pada *RapidMiner* adalah mengimpor data Excel ke dalam *read excel*.



Gambar 2. Upload Data

Selanjutnya adalah *setting* pada bagian *header*-nya dan mengubah *change role* yang terdapat pada *header* komentar menjadi label karena akan digunakan sebagai hasil klasifikasi dan mengubah *change type* pada *header* komentar menjadi *binominal*.

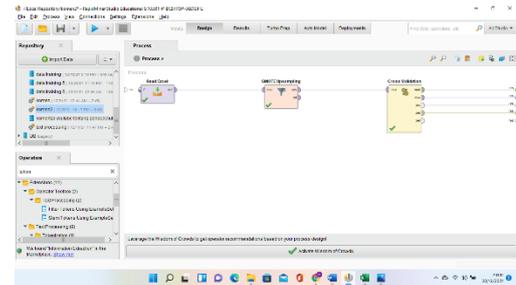


Gambar 3. Mengubah Data Label Pada Kolom Komentar

Pembuatan desain alur proses pada *RapidMiner* dilakukan dengan memasukkan proses tambahan seperti:

1. *Smote upsampling*, yang merupakan penyalaras antar kelas dengan membuat replikasi kelas minoritas sebanyak kelas mayoritas.
2. *Cross validation*, yang merupakan proses validasi data.

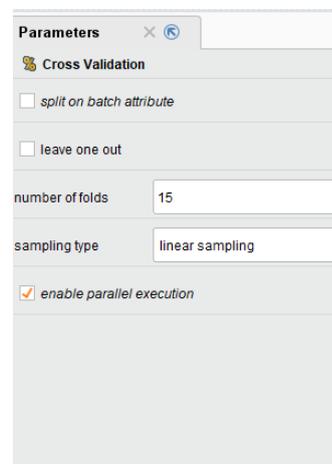
Berdasarkan penjelasan diatas dapat dilihat pada Gambar 4 yang merupakan desain *RapidMiner* klasifikasi *Naïve Bayes* yang digunakan.



Gambar 4. Desain RapidMiner Klasifikasi Naïve Bayes yang Digunakan

Proses Validasi

Proses yang dipakai pada *cross validation* yaitu operator *k-15 Fold cross validation* yang dapat dilihat pada konfigurasi parameter. Tampilan konfigurasi parameter validasi ini terdapat pada Gambar 5.



Gambar 5. Konfigurasi Parameter yang Digunakan

Di dalam *cross validation* terdapat proses yang digunakan yaitu:

- *Apply model* merupakan metode yang berguna untuk menyimpan data *training* dan data *testing*.
- *Naïve Bayes* digunakan untuk metode penghitungan data.
- *Performance* metode ini berguna untuk menghasilkan akurasi data yang biasanya dalam bentuk persentase.

Berdasarkan penjelasan di atas dapat dilihat pada Gambar 6 sebagai desain proses *cross validation*.

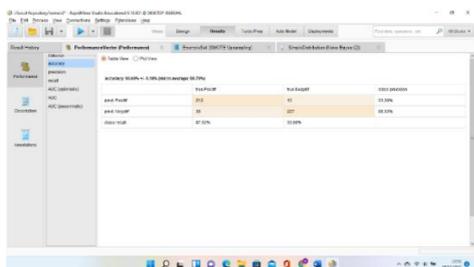


Gambar 6. Desain Proses Cross Validation

Evaluasi Dengan Smote dan Tanpa Smote

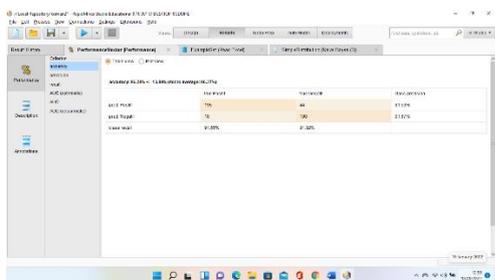
Evaluasi ini merupakan perbandingan dengan menggunakan smote dan tidak menggunakan smote dengan menganalisa sentimen berdasarkan *accuracy*

yang didapatkan. Hasil pengujian data *RapidMiner* dengan memakai *smote* mendapatkan *accuracy* sebesar 90.70%. Berikut bukti hasil *accuracy* dengan menggunakan *smote* yang terdapat pada Gambar 7.



Gambar 7. Hasil Analisis *Naive Bayes* Dengan Menggunakan *Smote*

Sedangkan hasil *RapidMiner* klasifikasi *Naive Bayes* tanpa memakai *smote* mendapatkan *accuracy* sebesar 86.37% seperti yang telah terlampir pada Gambar 8.



Gambar 8. Hasil Analisis *Naive Bayes* Tanpa Menggunakan *Smote*

Evaluasi Confusion Matrix

Selanjutnya penghitungan dengan *confusion matrix* yang diperoleh berdasarkan penelitian komentar masyarakat terhadap adanya *GeNose* yaitu kelas positif dan diklasifikasi dengan benar sebanyak 212. Hasil klasifikasi salah pada kelas positif adalah sebanyak 15 komentar. Hasil klasifikasi benar pada kelas negatif sebanyak 30 komentar, dan hasil klasifikasi salah pada kelas negatif sebanyak 227 komentar. Berdasarkan hasil tersebut, maka dapat dibuat tabel *confusion matrix* seperti pada Tabel 8.

Tabel 8. Hasil *Confusion Matrix*

	Positif	Negatif
Pred. Positif	212	15
Pred. Negatif	30	227

Berdasarkan tabel *confusion matrix*, dilakukan perhitungan nilai *recall*, *precision*, dan *accuracy* menggunakan rumus (2), (3), dan (4).

$$Recall = \frac{TP}{TP + FN} = \frac{212}{212 + 15} = 0.8760$$

$$Precision = \frac{TP}{FP + TP} = \frac{212}{15 + 212} = 0.9339$$

$$Accuracy = \frac{TP + TN}{Total} = \frac{212 + 227}{484} = 0.9070$$

Berdasarkan hasil perhitungan, dapat dilihat bahwa perhitungan *recall* untuk memprediksi komentar yang bersifat negatif mendapatkan persentase 87.60%. Nilai *precision*, yang digunakan untuk mengidentifikasi nilai prediksi positif pada kelas negatif berdasarkan keseluruhan hasil yang diprediksi negative, didapatkan persentase sebesar 93.39%. Nilai *accuracy* yang didapatkan pada penelitian ini sebesar 90.70%.

4. PENUTUP

Analisis sentimen tentang komentar masyarakat terhadap adanya *GeNose* dengan dataset pada komentar *Youtube* dan menggunakan metode *Naive Bayes* menghasilkan *accuracy* mencapai 90.70% dengan kategori komentar bersifat negatif. Nilai *recall* yang didapatkan sebesar 87.60%, sedangkan *precision* yang digunakan untuk mengidentifikasi prediksi positif pada kelas negatif mendapatkan persentase sebesar 93.39%. Data tersebut membuktikan bahwa *GeNose* kurang maksimal dalam meminimalisir *COVID-19*.

Saran yang dapat disampaikan untuk pembahasan adalah untuk mendapatkan hasil yang maskimal, maka dapat melakukannya dengan menggunakan data yang lebih baik lagi dengan menambahkan proses pada tahap *preprocessing*. Selain itu dapat memakai jumlah data yang lebih banyak dan juga dapat menambahkan metode lain untuk mendapatkan hasil yang lebih efisien.

REFERENSI

- [1] Simanjuntak, E Y., Oktavia, Y. T., & Sinaga, J. 2021. Optimalisasi Pencegahan Melalui Deteksi Dini Penularan Covid-19 Menggunakan *GeNose* C19. *Jurnal Pengabdian Masyarakat*, Vol. 1, No 2: 101-106.
- [2] Syahid, A. J., & Mahdiana, D. 2020. Algoritma Untuk Klasifikasi Analisis Sentimen Terhadap *Genose* Pada Media Sosial Twitter. *semanTIK*, Vol. 7, No. 1:1-5.
- [3] Permadi, V. A. 2020. Analisis Sentimen Menggunakan Algoritma *Naive Bayes* Terhadap Review Restoran Di Singapura. *Buana Informatika*, Vol.11, No.2:141-151.
- [4] Annur, H. 2018. Klasifikasi Masyarakat Miskin Menggunakan Metode *Naive Bayes*. *Ilkom Jurnal Ilmiah*, Vol. 10, No.2:160-165.
- [5] Putro, H. F., Vlandari, R. T., & Saptomo, W. L. Y. 2020. Penerapan Metode *Naive Bayes* Untuk Klasifikasi Pelanggan. *Jurnal Teknologi Informasi Dan Komunikasi (Tikomsin)*, Vol. 8, No.2:19-24.
- [6] Muin, A. 2016. Metode *Naive Bayes* Untuk Prediksi Kelulusan. *Jurnal Ilmiah Ilmu Komputer*, Vol.2, No.1:22-26.
- [7] Sulaksono, J., Heri Irawan, R., & Fahmi, I. N. Penerapan Metode *Naive Bayes* Terhadap Bantuan Sosial Keluarga Prasejahtera. *Nusantara of Engineering*, Vol. 3, No 2: 52-62.
- [8] S. Widaningsih & A. Suheri. 2018. Klasifikasi Jurnal Ilmu Komputer Berdasarkan Pembagian

- Web of Science Dengan Menggunakan Text Mining. Sentika, Hal 320-328, www.Elsevier.Com.
- [9] Chohan, S., Nugroho, A., Maezar Bayu Aji, A., Gata, W., & Mandiri, S. 2020. Analisis Sentimen Aplikasi Duolingo Menggunakan Metode Naïve Bayes Dan Synthetic Minority Over Sampling Technique. *Jurnal Informatika dan Komputer*, Vol. 22, No.2:139-144.
- [10] Dirjen, S. K. 2017. Klasifikasi Berita Menggunakan Algoritma Naive Bayes Classifier Dengan Seleksi Fitur Dan Boosting. *Jurnal Resti*, Vol. 1, No. 3: 227-232.
- [11] Hendrian, S. 2018. Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan. *Faktor Exacta*, Vol.11, No.3:266-274.
- [12] Watratan, A., Puspita, A. B., Moeis, D., Informasi, S., & Profesional Makassar, S. 2020. Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia. In *Journal of Applied Computer Science and Technology (Jacost)*, Vol. 1, No 1:7-14.
- [13] Yulia Hayuningtyas, R. 2019. Penerapan Algoritma Naïve Bayes Untuk Rekomendasi Pakaian Wanita. *Jurnal Informatika*, Vol. 6, No 1:18-22. [Http://Ejournal.Bsi.Ac.Id/Ejurnal/Index.Php/Ji/Article/View/4685](http://Ejournal.Bsi.Ac.Id/Ejurnal/Index.Php/Ji/Article/View/4685)
- [14] H. Irsyad, A. Farisi, And M. R. 2019. Klasifikasi Opini Masyarakat Terhadap Jasa Isp Myrepublic Dengan Naïve Bayes. *Jnteti*, Vol. 8, No.1:30-34.
- [15] Widhi Saputro, I., & Wulan Sari, B. 2019. Uji Performa Algoritma Naïve Bayes Untuk Prediksi Masa Studi Mahasiswa Naïve Bayes Algorithm Performance Test for Student Study Prediction. *Citec Journal*, Vol. 6 No 1:9-16.