

ANALISIS SENTIMEN KLASIFIKASI TWEET VAKSIN COVID 19 DENGAN NAÏVE BAYES

Blidex¹, Jati Sasongko Wibowo²

Teknik Informatika, Teknologi Informasi,
Universitas Stikubank Semarang, Universitas Stikubank Semarang
blidex09@gmail.com, jatisw@edu.unisbank.ac.id

ABSTRAK

Medsos Twitter memiliki pengguna sebanyak 332 juta dengan \pm 500 juta *tweet* yang dikirimkan per harinya dan 200 miliar *tweet* per tahun. Pemerintahan sekarang juga tidak lepas dari komentar publik tentang kebijakan vaksin Covid 19. Kebijakan Indonesia dalam program vaksinasi Covid 19 kebalikan dari kebijaksanaan sejumlah negara, para ahli mengatakan kelompok pertama yang divaksinasi haruslah staf medis garis depan dan kemudian orang tua. Pemerintah Indonesia lebih memilih untuk menyuntik vaksin Covid 19 bagi tenaga kesehatan yang berada di garis terdepan melawan pandemi Covid-19. Penelitian ini bertujuan untuk mengimplementasi algoritma *Naïve Bayes Classifier* (NBC) dalam mengklasifikasikan sebuah *tweet* opini sehingga dapat diketahui termasuk ke dalam kelas bersentimen positif atau negatif dan menganalisis nilai dari ketepatan algoritma *Naïve Bayes Classifier*. Metode NBC untuk menjalankan klasifikasi *tweet* vaksin Covid 19 dalam klasifikasi bersentimen positif atau negatif dengan 275 data latih dan 25 data uji yang diperoleh menghasilkan ketepatan 96 % dengan jumlah klasifikasi sentimen positif adalah 24 dan jumlah klasifikasi sentimen negatif adalah 0. Hasil yang diperoleh dalam proses pengklasifikasian mendapatkan akurasi yang cukup tinggi yaitu 96 % maka dari itu metode NBC bisa dipergunakan dalam menjalankan klasifikasi *tweet* vaksin Covid 19 positif atau negatif secara otomatis.

Kata Kunci: Twitter, Vaksin, Covid 19, Naïve Bayes.

PENDAHULUAN

Media sosial (medsos) merupakan tempat untuk berinteraksi antara pengguna dan ruang waktu didalam dunia digital. Medsos merupakan suatu bentuk yang terjadi di dunia nyata seperti plagiarisme (Nasrullah, 2016). Saat ini banyak medsos yang digunakan oleh masyarakat Indonesia diantaranya Facebook, Twitter, Instagram, Path, Line, dan masih banyak lagi. Medsos yang digunakan dalam penelitian ini adalah medsos Twitter.

Twitter merupakan wadah menuangkan suatu ide, gagasan atau tempat informasi yang dikumpulkan, pikiran yang diinspirasi, atau mengetahui pikiran dari teman [1]. Medsos Twitter memiliki pengguna sebanyak 332 juta dengan \pm 500 juta *tweet* yang dikirimkan per harinya dan 200 miliar *tweet* per tahun [2].

Perkembangan tren saat ini, pemerintah mencoba untuk bergerak lebih dekat ke model *citizen centric*, di mana prioritas dan jasa akan didorong sesuai dengan kebutuhan masyarakat daripada kemampuan pemerintah

[3]. Tren ini berkembang didasari oleh fenomena Arab Spring beberapa tahun yang lalu, bagaimana pemerintah dapat berdampak jika mengabaikan sentimen masyarakat. Tren tersebut memaksa pemerintah untuk berpikir ulang dan merancang kembali kebijakannya dalam berinteraksi dengan masyarakat. Untuk membangun pengetahuan tentang apa yang masyarakat inginkan bukanlah pekerjaan yang mudah, dengan jumlah informasi yang dihasilkan oleh medsos sangat besar. Untungnya, analisis sentimen dapat digunakan untuk mengatasi kasus tersebut.

Pemerintahan Indonesia juga melakukan pendekatan ke masyarakat menggunakan media sosial seperti Twitter. Hal ini didasari oleh himbuan Menpan tentang pemanfaatan media sosial untuk pemerintah [4]. Hasil survei PeerReach, Indonesia adalah pengguna Twitter terbanyak ke-3 di dunia sebanyak 6,5%. Peringkat pertama adalah Amerika

Serikat dengan jumlah 24,3% kemudian Jepang dengan jumlah 9,3%. Berdasarkan jumlah tersebut, diharapkan pemerintah Indonesia dapat menggali pengetahuan tentang sentimen masyarakat terhadap pemerintahan [5].

Pemerintahan sekarang juga tidak lepas dari komentar publik tentang kebijakan vaksin Covid 19. Kebijakan Indonesia dalam program vaksinasi Covid 19 kebalikan dari kebijaksanaan sejumlah negara, para ahli mengatakan kelompok pertama yang divaksinasi haruslah staf medis garis depan dan kemudian orang tua. Pemerintah Indonesia lebih memilih untuk menyuntik vaksin Covid 19 bagi tenaga kesehatan yang berada di garis terdepan melawan pandemi Covid-19 [6]. Pemanfaatan opini dari Twitter dibutuhkan analisis yang tepat sehingga informasi menggunakan analisis sentimen.

Analisis sentimen adalah proses pengolahan data tekstual secara otomatis untuk dipahami, diekstrak dari suatu opini ke dalam sentimen positif atau negatif menggunakan algoritma *Naïve Bayes Classifier* [7]. Metode *Naïve Bayes Classifier* (NBC) adalah metode klasifikasi untuk melakukan prediksi probabilitas dari suatu *class* [8].

Penelitian oleh Devita dkk [9] menyimpulkan metode NBC lebih baik dari metode KNN. Penelitian oleh Suryono dkk [10] menyimpulkan hasil klasifikasi NBC untuk sentiment positif = 28%, sentiment negatif = 20% dan sentiment netral = 52%. Penelitian oleh Rozi dkk [11], metode NBC menyatakan hasil uji akurasi untuk kategori positif = 82%, negatif = 92%, netral = 80 %.

LANDASAN TEORI

Text Mining

Text mining adalah penambangan yang dilakukan oleh komputer untuk didapatkan hal yang baru, hal yang belum diketahui sebelumnya secara implisit dari informasi yang diekstrak secara otomatis hasil sumber data teks yang berlainan [12]. *Text mining* dapat digunakan untuk kasus klasifikasi, *clustering*, *information extraction* dan *information retrieval* [13]. Dapat disimpulkan

text mining adalah sebuah sistem yang melakukan proses mengekstrak pengetahuan atau informasi yang dapat dimanfaatkan yang sebelumnya tidak diketahui.

Analisis Sentimen

Sentiment analysis (analisis sentimen) merujuk pada bidang yang luas dari penggunaan bahasa alami yang diolah, komputasi linguistik dan *text mining*. Terdapat banyak sebutan lain yang tugasnya sedikit berbeda yaitu *opinion mining*, *opinion extraction*, *sentiment mining*, dan lain-lain [7].

Twitter

Twitter adalah tempat menyalurkan berbagi ide, gagasan, informasi yang dikumpulkan, pikiran yang diinspirasi atau melihat kegiatan yang dilakukan oleh teman [14].

Naïve Bayes Classifier (NBC)

NBC merupakan algoritma untuk mencari nilai probabilitas tertinggi dari klasifikasi data uji dengan kategori yang paling tepat [8]. Pada saat klasifikasi, akan dicari probabilitas tertinggi dari semua kategori dokumen yang diujikan (VMAP), dimana persamaannya adalah

$$V_{map} = \underset{V_j \in V}{\operatorname{argmax}} \frac{P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j)}{P(x_1, x_2, x_3, \dots, x_n)}$$

Untuk $P(x_1, x_2, x_3, \dots, x_n)$ nilainya konstan untuk semua kategori (V_j) sehingga persamaan dapat ditulis persamaan berikut

$$V_{map} = \underset{V_j \in V}{\operatorname{argmax}} P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j)$$

Persamaan diatas dapat disederhanakan menjadi persamaan berikut

$$V_{map} = \underset{V_j \in V}{\operatorname{argmax}} \prod_{i=1}^n P(x_i | V_j) P(V_j) \quad (2.3)$$

Untuk $P(V_j)$ dan $P(x_i | V_j)$ dihitung pada saat pelatihan dimana persamaannya dapat dilihat di persamaan berikut

$$P(V_j) = \frac{|docs\ j|}{|contoh|}$$

$$P(x_i|V_j) = \frac{nk + 1}{n + kosakata}$$

METODE PENELITIAN

Metode Pengumpulan Data

Dalam penelitian ini, peneliti menggunakan berbagai metode untuk memperoleh data atau informasi saat memecahkan masalah untuk menganalisis sentimen tweet vaksin Covid 19. Metode yang digunakan antara lain:

1. Studi Kepustakaan

Metode pengumpulan data menggunakan metode kepustakaan untuk mengumpulkan data adalah dengan mengumpulkan terbitan berkala atau buku-buku sebagai sumber pustaka yang berkaitan dengan bahan tulisan, khususnya analisis sentimen menggunakan algoritma naive bayes classifier, seperti Prosiding Seminar Geotika Nasional, Jurnal Polinema Informatika, Jurnal Informasi pengembangan teknologi dan ilmu komputer, dan lain-lain.

2. Pengumpulan Data *Tweet*

Data yang dipergunakan adalah data yang bersumber dan dikumpulkan langsung dari web twitter yang berhubungan dengan vaksin Covid 19. Data *tweet* vaksin Covid 19 diperoleh dengan cara *crawling* (mengunduh secara otomatis) dengan menggunakan *spreadsheet online* dari google yang diperlihatkan seperti . gambar 3.1.



Gambar 3.1. *Spreadsheet Online*

Data *tweet* vaksin Covid 19 yang dihasilkan dari *spreadsheet online* dari google diperlihatkan seperti Tabel 3.1.

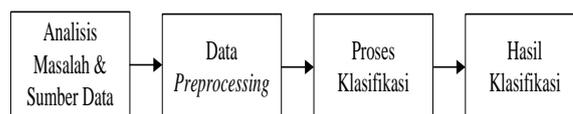
Tabel 3.1. Data *Tweet* Vaksin Covid 19

| No | Kategori | Text |
|----|----------|---|
| 1. | Positif | dalam percepatan memutuskan mata rantai penyebaran covid di kota padangsidimpuan kapolres |

| | | |
|-----|---------|---|
| | | padangsidimpuan hadiri rapat kordinasi terkait penambahan vaksin sinovac untuk kota padangsidimpuan bertempat di kantor dinas kesehatan kota padangsidimpuan rabu () |
| 2. | Positif | berita covid hasrat johor beli vaksin tambahan sudah dinyatakan kepada menteri luar hasni |
| 3. | Positif | terbaik pendaftaran penerimaan suntikan vaksin covid di indonesia menunjukkan peningkatan sejak awal bulan |
| .. | ... | .. |
| 300 | Positif | vaksin sinovac menjadi vaksin kedua yg diberikan ijin penggunaan oleh who ini semakin menegaskan bahwa vaksin sinovac memang aman digunakan untuk pencegahan covid |

Data *tweet* vaksin Covid 19 dari hasil *crawling* didapatkan data sebanyak 300 *tweet*. Data *tweet* vaksin Covid 19 kemudian dibagi menjadi 2 kelompok data yaitu data latih sebanyak 275 data *tweet* yang akan digunakan sebagai melatih model *Naive Bayes Classifier* dan data uji dibagi sebanyak 25 data *tweet* yang akan digunakan untuk menguji akurasi model *Naive Bayes Classifier*. Data *tweet* vaksin Covid 19 yang telah didapat akan disimpan dan disajikan dalam format Microsoft Excel xlsx.

Tahap Klasifikasi *Naive Bayes Classifier*



Gambar 3.3. Tahap Klasifikasi *Naive Bayes Classifier*

Algoritma analisis sentimen *tweet* vaksin Covid 19 menggunakan *Naive Bayes Classifier* diperlihatkan seperti gambar 3.3. Gambaran penyelesaian proses klasifikasi *tweet* vaksin Covid 19 menggunakan algoritma *Naive Bayes Classifier* sebagai berikut:

1. Analisis Masalah dan Sumber Data

Permasalahan yang terjadi dalam klasifikasi *tweet* vaksin Covid 19 yaitu:

- a. Banyak pengguna twitter yang melakukan posting ekspresi, opini atau pendapatnya terhadap vaksin Covid 19. Opini-opini tersebut dapat dimanfaatkan untuk mencari sebuah informasi.
- b. Belum adanya sistem yang dapat melakukan klasifikasi *tweet* vaksin Covid 19 positif atau negatif.

Untuk menanggulangi masalah *tweet* vaksin Covid 19 perlu dilakukan klasifikasi *tweet* vaksin Covid 19 menggunakan *text mining* dengan algoritma *Naïve Bayes Classifier*. Sumber data pada penelitian ini diambil secara langsung dari twitter yang berhubungan dengan vaksin Covid 19. Data yang telah didapat akan disimpan dan disajikan format Microsoft Excel *xlsx*. Setelah data didapatkan, langkah selanjutnya ialah melakukan penentuan klasifikasi secara manual terhadap data tersebut. Data yang diambil berjumlah 500 data *tweet* vaksin Covid 19 yang terdiri dari 450 *tweet* vaksin Covid 19 data latih dan 50 *tweet* vaksin Covid 19 data uji.

Data *tweet* vaksin Covid 19 akan melewati tahap *preprocessing* kemudian masuk ke tahap klasifikasi. Dalam penelitian ini, *tweet* vaksin Covid 19 akan diklasifikasikan menjadi dua klasifikasi yaitu klasifikasi *tweet* vaksin Covid 19 positif dan klasifikasi *tweet* vaksin Covid 19 negatif.

2. Data Preprocessing

Pada tahap ini, *tweet* vaksin Covid 19 berupa teks yang terkumpul akan melalui beberapa tahapan pemrosesan teks, termasuk case folding, tokenization, dan stop word removal. Tujuan dari pengolahan teks adalah agar data yang diperoleh lebih terstruktur, sehingga pengolahan data menjadi lebih mudah.

3. Proses Klasifikasi

Tahap selanjutnya adalah proses pengklasifikasian *tweet* vaksin Covid 19 yang akan diproses menggunakan metode algoritma *Naive Bayes Classifier* dengan menghitung probabilitas dari tiap *tweet* vaksin Covid 19.

4. Hasil Klasifikasi

Pada tahap ini adalah menentukan hasil dari klasifikasi menggunakan metode algoritma *Naive Bayes Classifier* dari proses klasifikasi, apakah *tweet* vaksin Covid 19 termasuk dalam klasifikasi *tweet* vaksin Covid 19 positif atau negatif.

HASIL DAN PEMBAHASAN

Membaca Data *Tweet* vaksin Covid 19

```
df <- read_excel("D:/Rstudio/blidex/data.xlsx")
glimpse(df)
```

Gambar 4.1 Membaca Data *Tweet* Excel

Gambar 4.1 merupakan *script* yang digunakan untuk memuat data *tweet* dalam format excel dengan tipe *.xlsx*. Dalam penelitian ini data yang digunakan adalah data excel yang berisi kumpulan data *tweet* vaksin Covid 19 positif atau negatif yang telah diberi kategori yang nantinya akan digunakan untuk proses pelatihan dan pengujian. Dalam penelitian ini data *tweet* vaksin Covid 19 positif atau negatif yang digunakan adalah data *tweet* vaksin Covid 19 positif atau negatif yang diambil dengan menggunakan teknik *crawling*.

Gambar 4.2 adalah tampilan dari hasil membaca data excel data.xlsx yang berisi 300 data *tweet* vaksin Covid 19 positif atau negatif dengan 2 variabel yaitu KATEGORI dan TEXT. Variabel KATEGORI berisi kategori *tweet* vaksin Covid 19 yaitu positif dan negatif, sedangkan variabel TEXT berisi kumpulan *tweet* vaksin Covid 19 positif atau negatif.

| KATEGORI | TEXT | |
|----------|---------|---|
| 1 | Positif | dalam percepatan memutus mata rantai penyebaran covi... |
| 2 | Positif | berita covid hasrat johor beli vaksin tambahan sudah di... |
| 3 | Positif | terbaik orang melaka trend pendaftaran penerimaan su... |
| 4 | Positif | dampak positif vaksinasi massal health immunity mulai t... |
| 5 | Positif | low pkp vaksin proklamasi darurat pejabat penerangan ... |
| 6 | Positif | berita covid warga emas gembira dapat suntikan vaksin ... |
| 7 | Positif | Jun ppv kuala terengganu capai sasaran kadar penerima... |
| 8 | Positif | vaksin pembudayaan norma baharu & proklamasi darur... |
| 9 | Negatif | proklamasi darurat vaksin covid task force pejabat pener... |
| 10 | Negatif | low pkpd vaksin covid pembudayaan norma baharu pej... |
| 11 | Negatif | takut untuk pergi vaksin sebab risau tak dapat perindu... |
| 12 | Positif | low pkpd proklamasi darurat nasional task force vaksin ... |
| 13 | Negatif | low vaksin pkp & proklamasi darurat pejabat peneranga... |
| 14 | Negatif | low vaksin pkp & proklamasi darurat pejabat peneranga... |

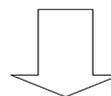
Gambar 4.2. Hasil Membaca Data *Tweet* Text Processing

Proses data *text preprocessing* yang dilakukan dalam penelitian ini yaitu

a. Case Folding

Pada tahap ini akan merubah semua huruf menjadi huruf kecil atau *lowercase*. Proses diawali dengan memeriksa setiap karakter *tweet* vaksin Covid 19 dari awal sampai akhir karakter, bila ditemukan huruf selain huruf kecil atau kapital, maka huruf yang dipilih akan diubah menjadi huruf kecil (*lowercase*). Proses *case folding* dapat dilihat seperti gambar 4.3.

Up dulu boi, vaksin bikin dompet tipis



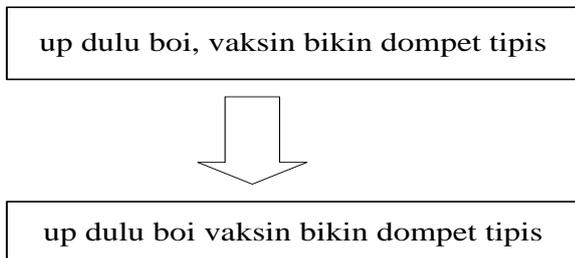
up dulu boi, vaksin bikin dompet tipis

Gambar 4.3. *Case Folding*

b. *Tokenizing*

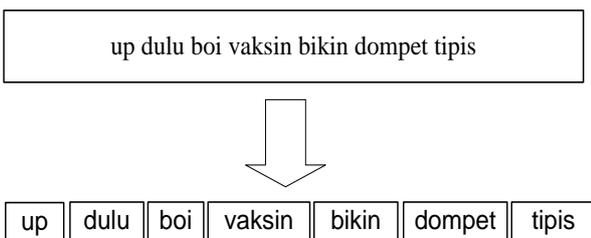
Pada tahap ini memotong sebuah *tweet* vaksin Covid 19 menjadi bagian-bagian, yang disebut dengan token dan pada saat bersamaan juga menghapus karakter khusus yang ada didalam *tweet* vaksin Covid 19, angka dan tanda baca diantaranya @, hashtag (#), titik (.), koma (,), dan tanda baca lainnya karena tidak memiliki pengaruh apapun terhadap nilai klasifikasi sehingga komponen-komponen tersebut harus dibuang.

Proses penghapusan karakter khusus yang ada didalam *tweet* vaksin Covid 19 diawali dengan memeriksa *tweet* vaksin Covid 19 dari proses *case folding* apakah karakter khusus, angka atau tanda baca, jika ada maka akan secara otomatis dihilangkan. Proses penghapusan karakter khusus dapat dilihat seperti gambar 4.4.



Gambar 4.4. Proses Penghapusan Karakter Khusus

Proses selanjutnya adalah memisahkan atau memotong *tweet* vaksin Covid 19 menjadi kata yang dipisahkan oleh tanda baca titik (.), koma (,), dan spasi. Proses diawali dengan mengecek *tweet* vaksin Covid 19 setelah di lakukan proses penghapusan karakter khusus, jika terdapat tanda baca titik (.), koma (,), dan spasi maka *tweet* vaksin Covid 19 akan dipisahkan atau dipotong berdasarkan tanda baca tersebut. Proses *tokenizing* dapat dilihat seperti gambar 4.5.

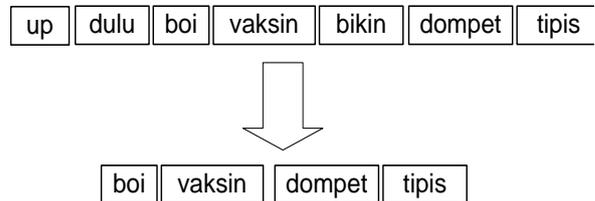


Gambar 4.5. *Tokenizing*

c. *Stopword Removal*

Setelah melewati tahap *tokenizing* maka tahap selanjutnya adalah tahap *stopword removal*. *Stopword* proses penghapusan kata yang bukan terlalu penting dari suatu dokumen. Misalnya “di”, “oleh”, “pada”, “sebuah”, “karena” dan lain-lain. Tahap *stopword removal* merupakan

pengerjaan untuk menghapus atau menghilangkan kata-kata umum seperti kata ganti, kata sambung, kata depan dan kata yang tidak penting akan dihilangkan. Proses diawali dengan melakukan pencocokan ke dalam database *stopword*, jika terdapat kata-kata yang ada pada database maka kata tersebut akan dihilangkan. Proses *stopword removal* dapat dilihat seperti gambar 4.6.

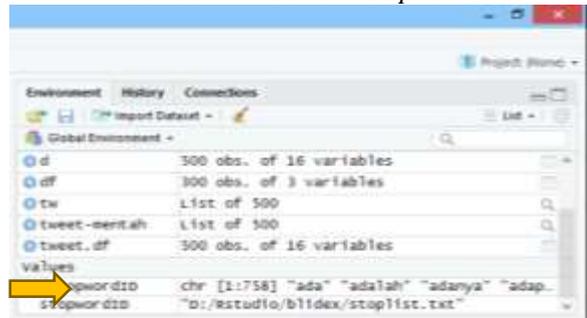


Gambar 4.6. *Stopword Removal*

RStudio belum menyediakan *stopword* dalam format bahasa Indonesia sehingga dibutuhkan *stopword* dari luar Rstudio untuk melakukan penghapusan kata sambung seperti gambar 4.7.

```
stopwordID <- "D:/Rstudio/blidex/stoplist.txt"
c$stopwordID<-readLines(stopwordID, warn=FALSE)
```

Gambar 4.7 *Custom Stopword*



Gambar 4.8. Hasil *Custom Stopword*

Gambar 4.8 menunjukkan *custom_stopwords* yang diimportkan memiliki 758 daftar kata. Daftar kata inilah yang akan dihilangkan jika ditemukan dalam *tweet* vaksin Covid 19 positif atau negatif. Untuk melaksanakan tahap *text processing* script bisa dilihat pada gambar 4.9.

```
corpus.clean <- tm_map(corpus,
content_transformer(tolower))
corpus.clean <- tm_map(corpus.clean,
removePunctuation)
corpus.clean <- tm_map(corpus.clean,
removeNumbers)
corpus.clean <- tm_map(corpus.clean,
removeWords,c$stopwordID)
corpus.clean <- tm_map(corpus.clean,
stripWhitespace)
```

Gambar 4.9. *Script Text Processing*

Hasil dari *text processing* terdiri dari beberapa proses yaitu *case folding*, *tokenizing* dan *stopword removal* disimpan dengan nama file *corpus.clean* yang bisa dilihat pada gambar dibawah ini.

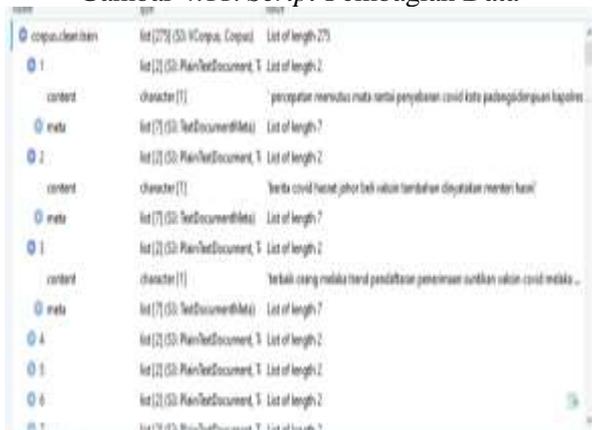


Gambar 4.10. Hasil *Text Processing*

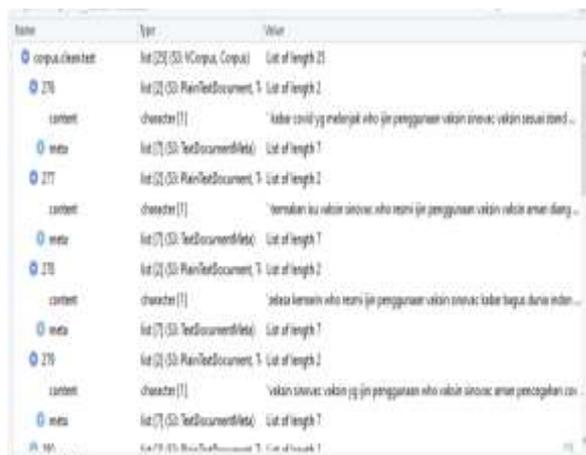
Data *tweet* vaksin Covid 19 positif atau negatif dipisah menjadi data uji dan data latih. Dari 300 data *tweet* vaksin Covid 19 positif atau negatif, data latih yang dipakai sebanyak 275 data *tweet* vaksin Covid 19 yang bisa dilihat di gambar 4.11, data uji berjumlah 25 data *tweet* vaksin Covid 19 yang dapat dilihat pada gambar 4.12. Data latih didapatkan 500 *term* yang bisa dilihat digambar 4.13. Setelah dilakukan *preprocessing*, *term* pada data latih berkurang menjadi 300 *term*.

```
show(dtm)
df.train <- df[1:275,]
df.test <- df[276:300,]
dtm.train <- dtm[1:275,]
dtm.test <- dtm[276:300,]
dim(dtm.train)
corpus.clean.train <- corpus.clean[1:275]
corpus.clean.test <- corpus.clean[276:300]
```

Gambar 4.11. *Script* Pembagian Data



Gambar 4.12. Data Latih



Gambar 4.13. Data Uji

| word | freq |
|------------------------|------|
| vaksin | 467 |
| covid | 367 |
| sinovac | 134 |
| who | 114 |
| penggunaan | 95 |
| jin | 79 |
| vaksinasi | 55 |
| resmi | 52 |
| aman | 51 |
| pengecahaan | 50 |
| kalbar | 37 |
| dun | 29 |
| low | 27 |
| ... | ... |
| to 14 of 1,425 entries | |

Gambar 4.14. *Term* Data Latih Hasil *Preprocessing*

Klasifikasi *Naive Bayes*

Untuk membentuk model data dipergunakan fungsi dari *Naive Bayes Classifier* yaitu *packages* 'e1071'. Karena *Naive Bayes Classifier* mengetes kemungkinan produk, maka diperlukan berbagai cara untuk memastikan kemungkinan non-nol ke kata-kata yang tidak terdpat pada spesimen dengan menggunakan *laplace = 1*. Gambar 4.15 merupakan *script* untuk model *Naive Bayes Classifier*.

```
trainNB <- apply(dtm.train.nb, 2, convert_count)
testNB <- apply(dtm.test.nb, 2, convert_count)
system.time( classifier <- naiveBayes(trainNB,
df.train$class, laplace = 1) )
predict(classifier, newdata=testNB)
predictions on the test set.
system.time( pred <- predict(classifier,
newdata=testNB) )
table("Predictions"= pred, "Actual" =
df.test$class )
conf.mat <- confusionMatrix(pred, df.test$class)
conf.mat
```

Gambar 4.15. Model *Naive Bayes Classifier*

```
[1] Positif Positif
[2] Positif Positif
[23] Positif Positif Positif
Levels: Negatif Positif
```

Gambar 4.16. Hasil Klasifikasi Data Uji

Gambar 4.16 merupakan hasil klasifikasi data uji menggunakan algoritma *Naive Bayes Classifier*.

Confusion Matrix (Menghitung Akurasi)

Proses perhitungan akurasi dari hasil klasifikasi *tweet* vaksin Covid 19 positif atau negatif menggunakan algoritma *Naive Bayes Classifier* pada data uji akan dicocokkan dengan sentimen yang telah diketahui sebelumnya.

| | Actual | |
|-------------|---------|---------|
| Predictions | Negatif | Positif |
| Negatif | 0 | 0 |
| Positif | 1 | 24 |

Gambar 4.17. Hasil *Confusion Matrix*

Dari gambar diatas menunjukkan data dari klasifikasi algoritma *naive bayes* dengan sentimen menghasilkan :

1. *True* Negatif menghasilkan 0 *tweet* vaksin Covid 19
2. *False* Negatif menghasilkan 1 *tweet* vaksin Covid 19
3. *False* Positif menghasilkan 0 *tweet* vaksin Covid 19
4. *True* Positif menghasilkan 24 *tweet* vaksin Covid 19

```
Accuracy : 0.96
95% CI : (0.7965, 0.999)
No Information Rate : 0.96
P-value [Acc > NIR] : 0.7358

Kappa : 0

McNemar's Test P-value : 1.0000

Sensitivity : 0.00
Specificity : 1.00
Pos Pred Value : NaN
Neg Pred Value : 0.96
Prevalence : 0.04
Detection Rate : 0.00
Detection Prevalence : 0.00
Balanced Accuracy : 0.50
```

Gambar 4.18. Akurasi Klasifikasi *Naive Bayes Classifier*

Pada gambar 4.18 akurasi dari algoritma *Naive Bayes Classifier* dalam menjalankan pengklasifikasian *tweet* vaksin Covid 19 positif atau negatif dalam penelitian ini sebesar 0.96 = 96%.

Pengukuran akurasi dilakukan dengan menggunakan *confusion matrix*. Data uji yang telah diklasifikasikan pada gambar 4.16 akan dibandingkan dengan data *tweet* vaksin Covid 19 yang sebenarnya. Hasil dari pengujian menggunakan *confusion matrix* menghasilkan

data matriks pada gambar 4.17 diperoleh 0 *tweet* vaksin Covid 19 negatif berhasil diklasifikasikan dengan benar, 1 *tweet* vaksin Covid 19 negatif salah diklasifikasikan ke dalam positif, 0 *tweet* vaksin Covid 19 positif salah diklasifikasikan ke dalam *tweet* vaksin Covid 19 negatif dan 24 *tweet* vaksin Covid 19 positif berhasil diklasifikasikan dengan benar. Dengan hasil dari *confusion matrix* maka didapatkan akurasi sebesar

$$\frac{TN+TP}{TN+TP+FN+FP} = \frac{24+0}{24+0+1+1} = \frac{24}{25} = 0,96 \dots\dots\dots$$

Hasil akurasi algoritma *Naive Bayes Classifier* sebesar 0,96 atau 96% dalam klasifikasi *tweet* vaksin Covid 19 positif atau negatif.

KESIMPULAN

Berdasarkan hasil pengujian dan penerapan yang telah dilaksanakan pada analisis sentimen twitter vaksin Covid 19 dengan *Naive Bayes Classifier*, maka akan diambil dengan kesimpulan seperti dibawah ini.

1. Metode *Naive Bayes Classifier* dalam melakukan klasifikasi *tweet* vaksin Covid 19 dalam klasifikasi bersentimen positif atau negatif dengan 275 data latih dan 25 data uji menghasilkan akurasi sebanyak 96 %.
2. Akurasi dari metode *Naive Bayes Classifier* memperlihatkan hasil sebanyak 96%, maka dari itu metode ini dapat digunakan untuk melakukan klasifikasi *tweet* vaksin Covid 19 positif atau negatif secara otomatis.

SARAN

Dilihat dari hasil penelitian yang telah diselesaikan, sistem yang dibangun masih memiliki kekurangan dari segi fungsi dan penyimpanan data. Oleh karena itu, untuk mendapatkan sistem yang lebih andal dan akurat, diperlukan pengembangan dan penyempurnaan lebih lanjut. Berikut ini adalah saran dari penulis agar sistem dapat berkerja dengan lebih baik dari penelitian ini yaitu :

1. Melakukan perbandingan dengan algoritma klasifikasi lain, untuk mengetahui algoritma yang memiliki akurasi terbaik.
2. Pengembangan ke dalam bahasa pemrograman lain.
3. Memperbanyak data *tweet* vaksin Covid 19 untuk mendapatkan hasil akurasi yang lebih bagus.

DAFTAR PUSTAKA

- [1] Nasrullah, R. *Media Sosial Perspektif Komunikasi, Budaya, Siosioteknologi*. Simbiosis Rekatama Media. Bandung. 2016
- [2] Maulana, A. *Twitter Rahasiakan Jumlah Pengguna di Indonesia*. Website: <https://www.cnnindonesia.com/teknologi/20160322085045-185-118939/twitter-rahasiakan-jumlah-pengguna-di-indonesia>. Diakses 20 September 2020. 2016
- [3] Arunachalam, R., & Sarkar, S. *The new eye of government: Citizen sentiment analysis in sosial media*. In Sixth International Joint Conference on Natural Language Processing (p. 23). 2015
- [4] Permenpan. *Pedoman Pemanfaatan Media Sosial Instansi Pemerintah*. 2012
- [5] Arifin, BH. *Pengguna Twitter Indonesia Terbanyak Ketiga Dunia*. <http://www.encycity.co/pengguna-twitter-indonesia-terbanyak-ketiga-unia>. Diakses 20 September 2021. 2015
- [6] A. Harun and D. P. Ananda, "Analisa Sentimen Opini Publik Tentang Vaksinasi Covid-19 di Indonesia Menggunakan Naïve Bayes dan Decision Tree," *Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. April, 2021.
- [7] Liu, B. *Sentiment Analysis And Opinion Mining*. Morgan dan Claypool Publisher. 2012
- [8] Santosa, B. *Data Mining: Teknik Pemanfaatan Data Untuk Keperluan Bisnis*. Graha Ilmu. Yogyakarta. 2015
- [9] Devita, R. N., Herwanto, H.W., & Wibawa, A. P., (2018) Perbandingan Kinerja Metode Naive Bayes Dan K-Nearest Neighbor Untuk Klasifikasi Artikel Berbahasa Indonesia. *Jurnal Teknologi Informasi dan Ilmu Komputer*. 5(4) September. pp. 427-434
- [10] Suryono, S., Utami, E., & Luthfi, E. T. Analisis Sentiment Pada Twitter Dengan Menggunakan Metode Naïve Bayes Classifier. *Prosiding Seminar Nasional Geotik*. pp.9-15. 2018
- [11] Rozi, I. F., Hamdana, E. N., & Alfahmi, M. B. I. Pengembangan Aplikasi Analisis Sentimen Twitter Menggunakan Metode Naïve Bayes Classifier (Studi Kasus SAMSAT Kota Malang). *Jurnal Informatika Polinema*. 4(2) Februari. pp.149-154. 2018
- [12] Symeonidis, S., Effrosynidis, D., & Arampatzis, A. A comparative evaluation of pre-processing techniques and their interactions for Twitter sentiment analysis. *Expert Systems With Applications*, 110, 298–310. 2018
- [13] Berry, M.W. & Kogan, J. *Text Mining Application and Theory*. Wiley. United Kingdom. 2015
- [14] Pratama, I. D. Bahasa Komplain di Media Sosial Twitter. *Jurnal Bahasa, Sastra, dan Pengajarannya*, 1(1), 35-56. 2017