

Analisis *Sentiment* Masyarakat terhadap Kasus Covid-19 pada Media Sosial Youtube dengan Metode *Naive bayes*

Muhammad Iqbal Ahmadi¹, Dudih Gustian², Falentino Sembiring³

^{1,2,3}Program Studi Sistem Informasi, Universitas Nusa Putra Sukabumi

Jl. Raya Cibolang No. 21 Sukabumi, West Java, Indonesia 43152

Muhammad.iqbal_si17@nusaputra.ac.id, Dudih@nusaputra.ac.id

Falentino.sembiring@nusaputra.ac.id

Abstract

The development of Covid-19 cases in Indonesia continues to increase. with the continued increase in these cases causing panic among the public regarding the presence or absence of this corona virus, in the midst of this condition an effective and efficient communication pattern is needed in providing education and information about this corona virus, for example with social media Youtube. Many people's responses to this news are expressed in the comments column. Therefore, a sentiment analysis model is needed to classify public comments into Positive, Negative and neutral. In this study, the Naive bayes method is used to build a sentiment analysis model for public responses about the news on the development of the Covid-19 case on the Youtube page, precisely on the KompasTV Chanel. Accuracy is 74% with the number of comments Positive 361, Negative 800 and neutral 490.

Keywords: Covid-19, sentiment analisis, Naive bayes, Youtube

Abstrak

Perkembangan kasus Covid-19 di Indonesia terus mengalami kenaikan. dengan terus adanya kenaikan kasus ini menyebabkan kepanikan di tangan masyarakat dengan ada dan tidaknya virus corona ini, di tengah kondisi ini di perlukan pola komunikasi yang efektif dan efisien dalam memberikan edukasi dan informasi tentang virus corona ini contohnya dengan media sosial Youtube. Banyak tanggapan masyarakat tentang pemberitaan ini yang di ungkapkan di kolom komentar. Oleh karena itu di butuhkan suatu model sentiment analisis untuk mengklasifikasikan komentar masyarakat menjadi Positif, Negatif dan netral. Dalam penelitian ini menggunakan metode Naive bayes dalam membangun model sentiment analisis tanggapan masyarakat tentang pemberitaan perkembangan kasus Covid-19 di laman Youtube tepatnya di Chanel KompasTV. Di dapatkan tingkat akurasi 74% dengan jumlah komentar Positif 361, Negatif 800 dan netral 490.

Kata kunci: Covid-19, sentiment analisis, Naive bayes, Youtube

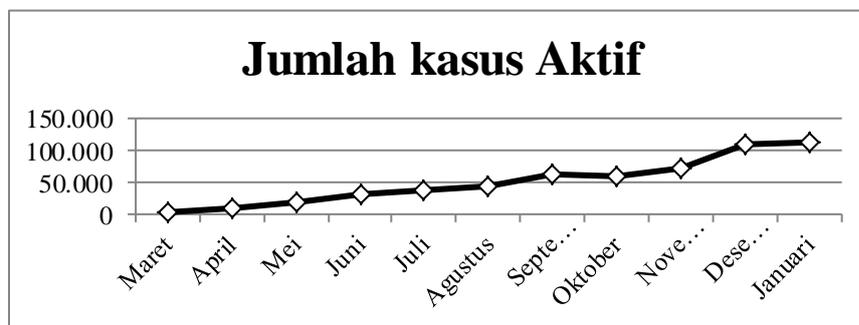
1. PENDAHULUAN

Perkembangan kasus COVID-19 di Indonesia dirasakan mengalami kenaikan yang terus menerus meningkat. berdasarkan laman COVID-19 nasional rata-rata kenaikan kasus aktif dari bulan April sampai Desember tahun 2020 sebesar 8.894 kasus. peningkatan ini dipicu oleh berbagai hal contohnya tingkat kepercayaan masyarakat terhadap adanya penyakit/virus COVID-19.

Tabel 1. Jumlah kasus aktif Covid-19 2020-2021

Bulan	Jumlah kasus Aktif
Maret	1.311

Bulan	Jumlah kasus Aktif
April	7.804
Mei	17.552
Juni	28.703
Juli	37.338
Agustus	41.329
September	61.321
Oktober	58.418
November	71.420
Desember	109.963
Januari	110.679



Gambar 1. Grafik bulanan kasus aktif Covid 19 2020-2021

Menurut laman DETIKHEALTH indonesia pertanggal 1 desember 2020 berada di peringkat 5 terbesar di asia dengan jumlah total kasus positif COVID-19 543.975 , meninggal 17.081, sembuh 454.879. diatas Banglades dengan jumlah kasus positif 467.225, meninggal 6.675, sembuh 383.224. dan di bawah Iraq dengan jumlah kasus 554.767, meninggal 12.306, sembuh 484.570 [1].

Di tengah kondisi ini di perlukan pola komunikasi yang efektif dan efisien untuk memberitakan edukasi dan informasi kepada publik tentang Covid-19. Di era digital ini banyak media sosial seperti Facebook, Twitter, instagram, whatsapp, dan youtube. Li HO-Y, Bailey A, Huynh D, et al. YouTube as a Source of Information On Covid-19: A Pandemic of Misinformation melalui jurnal BMJ Global Health 2020 menyatakan: Video berbahasa Inggris yang paling banyak dilihat terkait Covid-19 di YouTube, 27,5% atau sekitar 62 juta tayangan di seluruh dunia berisi informasi non faktual. Video-video tersebut berasal dari berita hiburan, berita internet, dan sumber konsumen[2]. Dalam penelitian tersebut hanya menggambarkan video berbahasa inggris saja, bisa di bayangkan berapa banyak jumlah video dengan berbagai bahasa dari seluruh dunia dan sangat banyak penonton di setiap harinya. oleh karena itu youtube di rasa menjadi sarana informasi yang sangat efektif dan efisien untuk komunikasi massa di tengah pandemi covid-19 ini.

Banyak media sosial youtube yang memberitakan tentang perkembangan kasus covid-19 di indonesia contohnya cnn indonesia, kompastv, cncb indonesia dan detik.com. namun sejauh ini kompastv mendapatkan jumlah penonton yang cukup tinggi, karena update berita yang hampir setiap hari dan data yang di ambil langsung dari laman satgas covid-19. Dengan terus adanya peningkatan jumlah aktif atau positif Covid-19 ini bisa di simpulkan bahwa masih banyaknya masyarakat yang tidak percaya dengan adanya virus Covid-19 di indonesia dan apa yang terus di beritakan di seluruh media sosial. Banyak sekali respon masyarakat terhadap pemberitaan Covid-19 ini banyak yang mempercayai pemberitaan tersebut, selain itu banyak juga masyarakat yang tidak percaya dan memberikan respon di kolom komentar di laman pemberitaan ini.

Dalam penelitian sebelumnya banyak algoritma yang di gunakan untuk sentiment analisis ini K-Nearest Neighbor (KNN) mendapatka nilai akurasi sebesar 58,10% dan algoritma *Naive bayes* memiliki tingkat akurasi yang lebih tinggi sebesar 63,21% [3]. Adapun penelitian lain menyatakan bahwa tingkat akurasi dari penggunaan algoritma seperti Support Vector Machine yang memiliki nilai akurasi 80,3% dan *Naive bayes* 80,81% [4]. Dalam penelitian lainnya algoritma *Naive bayes* juga menghasilkan tingkat akurasi yang cukup memuaskan dengan nilai akurasi sebesar 97,13%, dengan nilai precision 1 sementara pada class Recall dihasilkan nilai 95,49% dan nilai AUC 0,980 [5]. yang menggambarkan bahwa algoritma *Naive bayes* memiliki tingkat keberhasilan yang cukup menjanjikan.

Dalam penelitian Luthfika, dias Ramadhani dan annisa ferani yang berjudul Analisis Sentiment Masyarakat Terhadap COVID-19 pada Media Sosial Twitter menggunakan dua jenis algoritma yang digunakan dalam proses analisis sentiment yaitu K-Nearest Neighbor yang menghasilkan tingkat akurasi 82% dan *Naive bayes* dengan tingkat akurasi 85% yang menyimpulkan bahwa algoritma *Naive bayes* memberikan hasil yang baik dalam prosesnya [6].

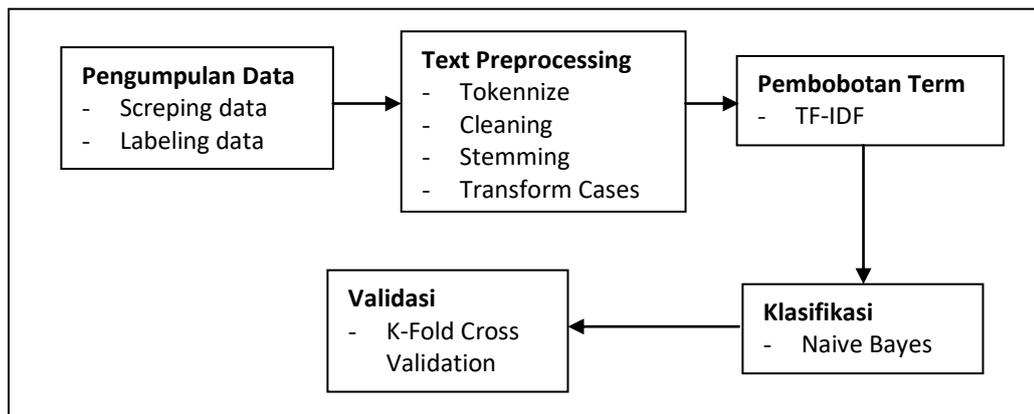
Pramana Yoga Saputra, Diah Hanifudin Subhi dan Fahmi Zain Afif Winatama dalam penelitiannya berjudul Implementasi Sentiment Analisis Komentar Chanel Video Pelayanan Pemeliharaan di Youtube Menggunakan Algoritma *Naive bayes*. Menjelaskan bahwa banyaknya opini yang di tuliskan di kolom komentar chanel video pelayanan pemerintah dapat dijadikan tolak ukur dalam menilai tingkat kepuasan masyarakat terhadap kinerja instansi dalam melaksanakan tugasnya. Oleh karena itu dalam penelitian ini kumpulan opini ini akan di bedakan menjadi 3 kategori yaitu positif, negatif dan netral, dan hasil dari pengujian tingkat akurasi dengan menggunakan metode *Naive bayes* menghasilkan 69,23% untuk KemenPUPR dan 64,10% untuk Kemenkeu[7].

Penelitian ini memberikan solusi untuk mengetahui tingkat persentase komentar dan respon masyarakat terhadap beberapa video yang memuat berita tentang perkembangan kasus COVID-19 di indonesia yang di terbitkan oleh chanel Kompastv, dan dapat meningkatkan tingkat kepercayaan

masyarakat terhadap adanya Virus COVID-19. Manfaat lain dari penelitian ini adalah mengukur kesuksesan pemerintah dalam menangani masalah Virus Covid-19. Dan mempresentasikan tingkat kepercayaan masyarakat terhadap berita yang di muat oleh channel Kompastv berdasarkan respon masyarakat di kolom komentar.

2. METODOLOGI PENELITIAN

Adapun tahap tahap pada penelitian ini antara lain :



Gambar 2. Metodologi Penelitian

2.1. Pengumpulan data

Dalam penelitian ini data di peroleh dari kolom komentar pemberitaan youtube di chanel KompasTV. Pengambilan data menggunakan tools yang di sediakan aplikasi *Google Crome* yang bernama Data miner.

2.2. Text Preprocessing

Tahap ini bisa juga di sebut tahap pembangunan data. Di tahap ini data akan memasuki beberapa proses diantaranya:

- Tokennize* yaitu tahap pemotongan string atau kalimat menjadi satuan kata.
- Cleaning* proses pembersihan karakter selain huruf, menghapus username atau mentions (@), hastag (#), dan menghapus ling atau URL dari setiap komentar.
- Stemming* merupakan proses mengubah kata berimbuhan menjadi kata dasar .
- Transform Cases* adalah tahap mengubah semua huruf kapital mejadi huruf kecil.

2.3. Labeling Data

Selanjutnya akan masuk ke tahap labeling yaitu tahap pengelompokan data menjadi 3 bagian yaitu Positif, Negatif, dan Netral berdasarkan nilai sekor sentiment. Sekor sentiment dihasilkan berdasarkan kamus sentiment positif dan negatif yang sudah banyak di gunakan[8].

2.4. Naive bayes

Naive bayes merupakan salah satu metode klasifikasi yang menggunakan perhitungan probabilitas, dalam metode ini menggunakan statistik berdasarkan teorema bayes yang mengasumsikan keberadaan dan ketiadaan dari suatu kelas dengan fitur lainnya[9].berikut adalah bentuk umum dari terema bayes [9]:

$$P(H|X) = \frac{p(X|H) \times p(H)}{p(X)} \quad (1)$$

$P(H|X)$ = Probabilitas hipotesis H berdasar kondisi X (posteriori probability)

$P(H)$ = Probabilitas hipotesis H (prior probability)

$P(X|H)$ = Probabilitas X berdasar kondisi pada hipotesis

$H P(X)$ = Probabilitas dari X

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan data

Data yang di ambil adalah sekumpulan data komentar dari beberapa video youtube pemberitaan perkembangan kasus Covid-19 di indonesia yang di publikasi di bulan januari yang di muat di chanel Kompastv berjumlah 28 buah file microsoft axcel, 1.651 komentar mentah.

3.2. Text Preprocessing

Pada tahap ini menggunakan aplikasi pendukung berupa Rstudio untuk mengolah data mentah. Adapun tahapan dari preprocessing diantaranya:

3.2.1. *Tokenize* adalah tahapan pemotongan string input berdasarkan kata, atau bisa disebut juga pemecahan kalimat menjadi potongan kata. Dalam tahap ini juga melakukan pembersihan karakter karakter tertentu seperti tanda baca.

3.2.2. *Cleaning proses* pembersihan karakter selain huruf, menghapus username atau mentions (@), hastag (#), dan menghapus link atau URL dari setiap komentar.

Tabel 2. *Cleaning data*

Hasil Tokenize	Hasil Cleaning
Tidak percaya LAGI. Tidak akan aku Mempercayai. CORONA.	Tidak Percaya Tidak Mempercayai CORONA

3.2.3. *Stemming* merupakan proses mengubah kata berimbuhan menjadi kata dasar .

Tabel 3. Steaming Data

Hasil <i>Cleaning</i>	Hasil <i>Stemming</i>
Tidak	Tidak
Percaya	percaya
Tidak	Tidak
Mempercayai	Percaya
CORONA	CORONA

3.2.4. *Transform Cases* adalah tahap mengubah semua huruf kapital mejadi huruf kecil

Tabel 4. Transform Cases

Hasil <i>Stemming</i>	Hasil <i>Transform case</i>
Tidak	Tidak
percaya	percaya
Tidak	tidak
Percaya	percaya
CORONA	corona

3.3. Labeling Data

Data yang sudah siap selanjutnya akan masuk ke tahap labeling yaitu tahap pengelompokan data menjadi 3 bagian yaitu Positif, Negatif, dan Netral berdasarkan nilai skor sentiment. Skor sentiment dihasilkan berdasarkan kamus sentiment positif dan negatif yang sudah banyak di gunakan[8].

Tabel 5. hasil labeling data

Negatif	Positif	Netral
800	361	490

score	text	klasifikasi
-2	akhir akhir ini semakin sering ngelihat ambulan ber apd len...	Negatif
0	perlawanan resistance	Netral
0	vaksin kapan distribusi om	Netral
0	tambahe kok sansoyo mbudak piye iki	Netral
1	alhamdulillah yg sembuh semakin banyak ya allah	Positif
-1	tetangga meninggal kma sakit jantung di tawari di fasilitas ...	Negatif
0	alhamdulillahh yg sembuh tambah banyak mudah mudaha...	Netral

Gambar 3. Hasil Labeling Data

3.4. Validasi *K-Fold Cross Validation*

Pada tahap Validasi ini bertujuan untuk mengukur dan mengevaluasi kinerja sebuah algoritma. Di tahap ini juga data akan di bagi menjadi dua

subset yaitu data *Training* dan data *Testing*, namun pada tahap ini hanya melibatkan dua data set saja yaitu data yang berlabel Positif dan Negatif sedangkan data yang berlabel Netral tidak diikutsertakan karena akan lebih fokus ke komentar positif dan negatif saja. Pembagian data ini menggunakan perbandingan 1 : 3 .pada data *Training* diberikan 75% dari keseluruhan data yaitu berjumlah 871 data. Seangakan untuk data *Testing* diberikan persentase data 25% data dari keseluruhan data sebanyak 290 dari keseluruhan data.

```
inspect(corpus[1:3])

dtm <- DocumentTermMatrix(corpus)
inspect(dtm[1:15, 10:15])

df.train <- df[1:871,]
df.test <- df[872:1161,]

dtm.train <- dtm[1:871,]
dtm.test <- dtm[872:1161,]

corpus.train <- corpus[1:871]
corpus.test <- corpus[872:1161]
```

Gambar 4. pembagian data *Testing* dan data *Training*

Pada tahap ini juga data akan di bagi menjadi 10 Fold yang di mana data akan dilakukan randomasi data agar tidak menjadi pengelompokan data.

3.5. Akurasi

Setelah melalua seluruh tahapan yang ada dari pengambilan data, text Preprocessing, labeling data, dan yang terakhir tahap validasi, didapatkan nilai akurasi dari algoritma yang di gunakan yaitu *Naive bayes* menghasilkan tingkat akurasi yang cukup baik yaitu sebesar 74%.

```
> conf.mat$byClass
  Sensitivity      Specificity    Pos Pred value    Neg Pred value      Precision
0.5517241      0.8735632      0.7441860      0.7450980      0.7441860
  Recall
0.5517241
  F1
0.6336634
  Prevalence
0.4000000
  Detection Rate
0.2206897
  Detection Prevalence
0.2965517
  Balanced Accuracy
0.7126437

> conf.mat$overall
  Accuracy      Kappa  AccuracyLower  AccuracyUpper  AccuracyNull  AccuracyPValue  McNemarPValue
7.448276e-01  4.444444e-01  6.905788e-01  7.939968e-01  6.000000e-01  1.571740e-07  7.484652e-04
> conf.mat$overall['Accuracy']
Accuracy
0.7448276
```

Gambar 5. hasil tingkat akurasi metode *Naive bayes*

4. SIMPULAN

Dari seluruh tahap penelitian yang sudah di lakukan dapat di simpulkan bahwa tanggapan masyarakat lebih condong ke sisi negatif. dengan jumlah komentar negatif sebanyak 800 komentar dan jumlah komentar positif yang di berikan masyarakat sebanyak 361 komentar. Kesimpulan lain dari

penelitian ini juga bahwa kinerja dari metode algoritma *Naive bayes* berjalan dengan cukup memuaskan dengan menghasilkan tingkat akurasi sebesar 74%.

DAFTAR PUSTAKA

- [1] satgas Covid-19, "Analisis Data Covid-19 Indonesia Update Per 03 Januari 2021." https://covid19.go.id/storage/app/media/Analisis Data COVID-19 Indonesia/2021/Januari/Analisis Data COVID-19 Mingguan Satuan Tugas PC19 per 03 Januari 2021 vFinal_compressed.pdf (accessed Jun. 08, 2021).
- [2] H. O. Y. Li, A. Bailey, D. Huynh, and J. Chan, "YouTube as a source of information on COVID-19: A pandemic of misinformation?," *BMJ Glob. Heal.*, vol. 5, no. 5, May 2020, doi: 10.1136/bmjgh-2020-002604.
- [3] M. Syarifuddin, "Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naïve Bayes Dan Knn," *Inti Nusa Mandiri*, vol. 15, no. 1, pp. 23–28, 2020.
- [4] Sharazita Dyah Anggita and Ikma, "Algorithm Comparison of *Naive bayes* and Support Vector Machine based on Particle Swarm Optimization in Sentiment Analysis of Freight Forwarding Services," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 2, pp. 362–369, 2020, doi: 10.29207/resti.v4i2.1840.
- [5] D. G. Rita apriani, "Analisis Sentimen Dengan Naïve Bayes Terhadap Komentar Aplikasi Tokopedia," vol. 6, no. 1, 2019.
- [6] A. L. Fairuz, R. D. Ramadhani, and N. A. Tanjung, "Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial," *J. DINDA*, vol. 1, no. 1, pp. 10–12, 2021, [Online]. Available: <http://journal.ittelkom-pwt.ac.id/index.php/dinda/article/view/180>.
- [7] P. Y. Saputra, D. H. Subhi, and F. Z. A. Winatama, "Implementasi Sentimen Analisis Komentar Channel Video Pelayanan Pemerintah Di Youtube Menggunakan Algoritma Naïve Bayes," *J. Inform. Polinema*, vol. 5, no. 4, pp. 209–213, 2019, doi: 10.33795/jip.v5i4.259.
- [8] W. Anggraini, M. Utami, J. Berlianty, and E. Sellya, "Klasifikasi Sentimen Masyarakat Terhadap Kebijakan Kartu Prakerja di Indonesia," *Fakt. Exacta*, vol. 13, no. 4, pp. 255–261, 2021, doi: 10.30998/faktorexacta.v13i4.7964.
- [9] M. Christianto, J. Andjarwirawan, and A. Tjondrowiguno, "Aplikasi analisa sentimen pada komentar berbahasa Indonesia dalam objek video di website YouTube menggunakan metode Naïve Bayes classifier," *J. Infra*, vol. 8.1, pp. 255–259, 2020.