

IMPLEMENTATION OF THE NAIVE BAYES CLASSIFIER ALGORITHM FOR CLASSIFICATION OF COMMUNITY SENTIMENT ABOUT DEPRESSION ON YOUTUBE

Sri Mulyani^{*1}, Rice Novita²

^{1,2}Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, UIN Suska Riau, Indonesia
Email: ¹11753201985@students.uin-suska.ac.id, ²rice.novita@uin-suska.ac.id

(Naskah masuk: 10 Juni 2022, Revisi: 25 Juni 2022, diterbitkan: 24 Oktober 2022)

Abstract

Depression is a disease that knows no age, gender and social status. WHO states that more than 264 million people suffer from depression, people with depression will continue to grow if public knowledge about mental health is still low, especially in Indonesia. This can be known from the way the community responds to a case. This study aims to determine public sentiment towards people with depression by classifying comments using the Naive Bayes Classifier (NBC) algorithm and adding the Term Frequency-inverse Document Frequency (TF-IDF) method as a feature extraction method. Sentiment used as data is obtained from YouTube comments on several news media accounts such as tvOneNews, Kompas TV, Tribunnews, Official iNews, VIVACOID, CNN Indonesia and Tribun Jateng, so that 4783 data are obtained with training data of 3826 and 957 testing data. This sentiment was analyzed by giving three classes, namely positive, neutral and negative. The results of the sentiment analysis were dominated by positive sentiment of 93.31%, followed by negative comments of 6.68% while neutral sentiment was 0%, and the accuracy of the NBC Algorithm was 84.11%.

Keywords: *Depression, NBC, TF-IDF, YouTube.*

IMPLEMENTASI ALGORITMA NAIVE BAYES CLASSIFIER UNTUK KLASIFIKASI SENTIMEN MASYARAKAT TENTANG DEPRESI PADA YOUTUBE

Abstrak

Depresi merupakan penyakit yang tidak mengenal usia, jenis kelamin dan status sosial. WHO menyatakan lebih dari 264 juta jiwa menderita depresi, penderita depresi akan terus bertambah jika pengetahuan masyarakat tentang kesehatan mental masih rendah terutama di Indonesia. Hal ini dapat diketahui dari cara masyarakat menanggapi suatu kasus. Penelitian ini bertujuan untuk mengetahui sentimen masyarakat terhadap penderita depresi dengan mengklasifikasikan komentar menggunakan algoritma *Naive Bayes Classifier* (NBC) dan ditambahkan metode Term Frequency-inverse Document Frequency (TF-IDF) sebagai metode ekstraksi fitur. Sentimen yang digunakan sebagai data di dapat dari komentar YouTube di beberapa akun media berita seperti tvOneNews, Kompas TV, Tribunnews, Official iNews, VIVACOID, CNN Indonesia dan Tribun Jateng, sehingga didapatkan data sebanyak 4783 dengan data *training* sebesar 3826 dan 957 data *testing*. Sentimen ini dianalisis dengan memberikan tiga kelas yaitu positif, netral dan negatif. Hasil analisis sentimen didominasi oleh sentimen positif sebesar 93,31% kemudian disusul oleh komentar negatif sebanyak 6,68% sedangkan sentimen netral 0%, dan didapatkan keakurasian Algoritma NBC sebesar 84,11%.

Kata kunci: *Depresi, NBC, TF-IDF, YouTube.*

1. PENDAHULUAN

Kesehatan mental sedang menjadi topik yang sering dibicarakan terutama pada sosial media, salah satunya depresi. Depresi merupakan suatu kondisi psikis seseorang yang sedang menurun sehingga mengakibatkan sedih mendalam, tidak peduli terhadap lingkungan, stress, sakit. Gejala yang timbul jika terus diabaikan dapat mengganggu aktivitas [1] bahkan membawa penderita melakukan hal diluar

batasan sehingga memilih kematian. Gejala ini menunjukkan, bahwa depresi bukan lah hal yang dapat disepelekan [2].

Organisasi kesehatan dunia (WHO) menyatakan, lebih dari 264 juta jiwa dari berbagai kalangan usia menderita depresi. Penyakit ini tidak memandang jenis kelamin, ras, usia dan status sosial [3]. Jumlah penderita terus meningkat, rehabilitasi menjadi cara yang dilakuakn untuk menekan laju

pertambahannya penderita depresi. Upaya ini tidak akan berhasil jika masyarakat masih melakukan intimidasi terhadap penderita yang berada di sekitar mereka [4]. hal ini terjadi karena pemahanan yang rendah tentang kesehatan mental dan mengakibatkan terhambatnya penyembuhan dikarenakan penderita takut untuk mendapatkan bantuan profesional dan akhirnya penderita dapat menyakiti diri sendiri atau menyakiti orang lain.

Seperti kasus “seorang ibu tega gorok 3 anak di brebes” yang terjadi pada 20 Maret 2022, diberitakan bahwa pada kejadian ini seorang ibu berinisial KU(35thn) menggorok 3 orang anaknya, dengan 1 anak tewas dan 2 lainnya dapat diselamatkan, yang langsung mendapatkan perawatan di sebuah rumah sakit. Setelah melakukan pendalaman pada kasus ini dan observasi terhadap si ibu, kepolisian brebes menyatakan bahwa si ibu mengalami depresi berat. Kasus ini diangkat oleh beberapa media seperti tvOneNews, Kompas TV, Tribunnews, Official iNews, VIVACOID, CNN Indonesia dan Tribun Jateng yang diunggah pada sosial media berupa youtube. Menurut *Pew Research Center* pada tahun 2021 YouTube menjadi media sosial paling banyak digunakan, YouTube unggul dari beberapa media lain seperti Facebook, Instagram, Snapchat, Twitter dan Tiktok [5]. YouTube menarik pengguna dengan memanfaatkan kekurangan Televisi(TV), yaitu dengan iklan yang sedikit, durasi video yang bervariasi dan YouTube menyajikan berbagai video yang dapat di pilih pengguna [6]. Dengan adanya aplikasi ini, pemberian informasi dapat menyebar dengan lebih luas dan cepat, dan pengguna dapat mengekspresikan pendapatnya dalam bentuk komentar [7] Pada kolom komentar ini masyarakat dapat dengan bebas berpendapat. Pada penelitian ini didapatkan data komentar sebanyak 4783 data komentar.

Komentar yang berisi sentimen terhadap kasus tersebut dapat dilakukan sebuah penggalan informasi berupa pengklasifikasian opini positif, netral ataupun negatif [7]. Penggalan informasi dapat dilakukan dengan menggunakan metode *Text Mining*. Metode ini sendiri berarti penambangan data teks dimana data yang digunakan berasal dari berbagai dokumen dan memiliki tujuan untuk mencari kata-kata yang mampu mewakili isi dokumen sehingga dapat dilakukan analisis keterikatan antar dokumen [8].

Text Mining dilakukan dengan ditambahkan sebuah algoritma klasifikasi, pada penelitian ini digunakan algoritma *Naive Bayes Classifier* (NBC). Dan kelebihan NBC berupa cepat dalam perhitungan, algoritma yang sederhana namun memiliki keakurasian yang tinggi dan mampu menangani data yang tidak lengkap [9]. Untuk membuktikan keunggulan NBC dalam hal klasifikasi terhadap sentimen, telah dilakukan penelitian yang membandingkan NBC dan *Support Vector Machine* (SVM) dengan judul penelitian “Analisis sentimen calon gubernur DKI Jakarta 2017 di Twitter” dengan

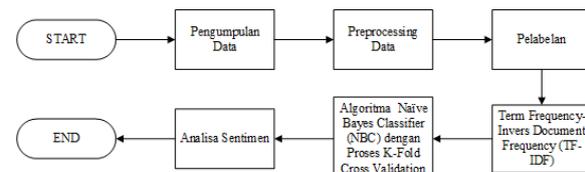
hasil akurasi NBC sebesar 95% [10]. Pada penelitian sentimen analisis lainnya algoritma NBC di sandingkan dengan dua algoritma lain seperti *Decision Tree* dan *Random Forest Algorithm*, dan pada penelitian kali ini pun NBC masih unggul dengan akurasi 84.43% [11].

Algoritma NBC diaplikasikan pada data komentar yang telah di kumpulkan, pada penelitian ini ditambahkan metode *Term Frequency-inverse Document Frequency* (TF-IDF). Metode TF-IDF ditambahkan karena mampu meningkatkan akurasi pada algoritma klasifikasi terutama NBC [12]. TF-IDF digunakan karena selain meningkatkan keakurasian metode ini juga merupakan metode yang mudah untuk diterapkan [13].

Proses pengerjaan pada penelitian ini dilakukan menggunakan bahasa pemrograman python, yang nantinya hasil yang di dapat berupa akurasi algoritma dan analisis sentimen yang telah di kelompokkan berdasarkan kelas negatif, netral, dan positif.

2. METODE PENELITIAN

Secara garis besar, ada enam tahapan yang dilakukan dalam penelitian ini. Berikut tahap-tahap dalam penelitian dapat dilihat pada gambar 1.



Gambar 1. Metodologi Penelitian

Tahapan awal ialah pada penelitian ini ialah pengumpulan data, data yang dikumpulkan merupakan data komentar yang di dapat dengan teknik *scrapping*. Data yang di *scrapping* merupakan data komentar dari semua video YouTube dengan pembahasan kasus “seorang ibu tega gorok 3 anak di brebes”, dan yang berasal dari akun media seperti tvOneNews, Kompas TV, Tribunnews, Official iNews, VIVACOID, CNN Indonesia dan Tribun Jateng. Jumlah komentar yang di kumpulkan sebanyak 4783 komentar, yang nantinya akan digunakan sebagai data penelitian. Selanjutnya akan dilakukan *preprocessing data* untuk mengubah data menjadi terstruktur. Setelah dilakukannya *preprocessing*, data yang dikumpulkan akan dilakukan pelabelan untuk menentukan sentimen, pada penelitian ini pelabelan dilakukan secara manual oleh pakar bahasa yang bernama Nuratika, S.Hum, M.Pd. Data yang sudah diberi label sentimen selanjutnya akan dilakukan pembobotan kata menggunakan *Term Frequency-Invers Document Frequency* (TF-IDF). Selanjutnya akan diterapkan algoritma NBC dan tahapan akhir pada penelitian ini adalah melakukan analisis hasil klasifikasi NBC, analisis sentimen dan visualisasi kata.

2.1. Text Mining

Text mining ialah suatu proses untuk mengekstraksi pola dari dokumen yang tak terstruktur dan nantinya akan didapatkan sebuah pola menarik yang dapat digunakan[14]. Tujuan utama metode ini adalah menemukan kata-kata yang mewakili dokumen.

2.2. Term Frequency-Invers Document Frequency (TF-IDF)

Dataset yang telah melewati proses *preprocessing* masih berbentuk kata-kata, dan harus di ubah kedalam bentuk numerik pada penelitian ini pembobotan dilakukan dengan menggunakan TF-IDF. Metode ini bekerja untuk menentukan keterikatan kata (*term*) terhadap dokumen dengan memberikan nilai pada setiap kata[15]. Tujuan metode ini ialah mencari kata-kata yang dapat menunjukkan dokumen teks pada suatu koleksi. TF-IDF menjadi metode terbaik untuk mendapatkan informasi[16]. Metode ini memiliki rumus sebagai berikut[17] :

$$W_{x,n} = TfIdf_{x,n} \cdot a^{w_{x,n-1}} \quad (1)$$

Keterangan :

$W_{x,n}$ = bobot kata x pada interval ke-n

$TfIdf_{x,n}$ = Nilai Tf-Idf kata x pada interval ke-n

a = *Discounted cumulative factor*

2.3. Naive Bayes Classifier (NBC)

Naive Bayes ditemukan oleh seorang ilmuwan yang bernama Thomas Bayes pada tahun 1763. NBC memanfaatkan data *sample* yang diperoleh dari populasi dan memperhitungkan suatu distribusi awal yang disebut distribusi *prior*. Probabilitas *bayesian* adalah suatu cara untuk mengatasi ketidakpastian dengan menggunakan formula *bayes*[18] yang dinyatakan sebagai berikut:

$$P(H | E) = \frac{p(E | H) \cdot p(H)}{p(E)} \quad (1)$$

Keterangan:

$P(H | E)$ = peluang hipotesa H jika terdapat *evidence* E

$P(E | H)$ = peluang munculnya *evidence* E jika hipotesis H

$P(H)$ = peluang hipotesa H tanpa memandang *evidence* apapun

$P(E)$ = peluang *evidence* E tanpa memandang apapun

3. HASIL DAN PEMBAHASAN

Data awal yang akan digunakan pada penelitian ini berjumlah 4783 data komentar yang dikumpulkan dengan metode *scrapping* menggunakan bahasa pemrograman *Phyton* pada beberapa akun Youtube media berita. Selanjutnya data yang telah

dikumpulkan akan melewati tahapan *preprocessing* sebelum diterapkan algoritma NBC, karena data masih berbentuk kata-kata dan harus berbentuk numerik, maka data harus diubah dengan cara memberikan bobot kata menggunakan metode TF-IDF. Dan selanjutnya *dataset* dapat digunakan untuk pengujian algoritma NBC.

3.1. Data Awal

Pada tahapan ini, data penelitian yang digunakan adalah data komentar yang didapatkan dengan cara *scrapping* menggunakan bahasa pemrograman *Phyton* pada akun YouTube beberapa akun media berita. Berikut adalah data awal yang digunakan, dapat dilihat pada tabel 1.

Tabel 1. Data Awal

No	Username	Komentar
1	Mengwibu	Gk paham lagi, indo apa ² masalahnya pertumpahan darah di jadikan solusi, cekcok lgsg adu bacok, ribut soal utang adu bacok. Protes pedagang buar di pasar pembeli tewas dibacok, cari makan sahur tewas jadi korban klitih, ibu saksikan anaknya tewas di keroyok pake sajam, acara nikahan jadi acara pembantaian, nyawa orang kek gk ada harganya di negara ini
2	mello coot	Ya allah
3	Rahma Wati	Anak adalah titipan allah yang harus kita jada dn rawat dgn benar.
...
4788	Ashanty Nurjayanti	Streeess itu buibu

3.2. Preprocessing

Tahapan ini dilakukan untuk menyiapkan data awal sebelum dilakukanya proses klasifikasi menggunakan algoritma. Data yang berasal dari internet merupakan data tak terstruktur (*unsupervised*) maka dibutuhkan suatu proses yaitu proses *preprocessing* agar data tersebut dapat dignakan

3.2.1. Case Folding

Case folding dilakukan untuk menghapus *username* dan mengubah semua karakter huruf menjadi huruf kecil hasil dari proses *case folding* dapat dilihat pada Tabel 2.

Tabel 2. Hasil *Case folding*

No	Komentar
1	gk paham lagi, indo apa ² masalahnya pertumpahan darah di jadikan solusi, cekcok lgsg adu bacok, ribut soal utang adu bacok. protes pedagang buar di pasar pembeli tewas dibacok, cari makan sahur tewas jadi korban klitih, ibu saksikan anaknya tewas di keroyok pake sajam, acara nikahan jadi acara pembantaian, nyawa orang kek gk ada harganya di negara ini
2	ya allah
3	anak adalah titipan allah yang harus kita jada dn rawat dgn benar.

...
4783 streeess itu buibu

3.2.2. Cleaning

Tahapan *cleaning* merupakan tahapan penghapusan angka, tanda baca (*delimiter*), penggunaan *uniform resource locator* (URL) dan penghapusan *hashtag* (#). Hal ini dikarenakan, pengguna yang berkomentar masih banyak yang menggunakan simbol-simbol tersebut. Hasil dari proses *cleaning* dapat dilihat pada Tabel 3.

Tabel 3. Hasil *Cleaning*

No	Komentar
1	gk paham lagi, indo apa ² masalahnya pertumpahan darah di jadikan solusi, cekcok lgsg adu bacok, ribut soal utang adu bacok. protes pedagang buar di pasar pembeli tewas dibacok, cari makan sahur tewas jadi korban klitih, ibu saksikan anaknya tewas di keroyok pake sajam, acara nikahan jadi acara pembantaian, nyawa orang kek gk ada harganya di negara ini
2	ya allah
3	anak adalah titipan allah yang harus kita jada dn rawat dgn benar.
...	...
4783	streeess itu buibu

Setelah menambahkan *stoplist* maka akan didapatkan hasil *filtering* yang dapat dilihat pada Tabel 4.

Tabel 4. Hasil *Filtering*

No	Komentar / <i>filtering</i>
1	paham indo pertumpahan darah jadikan solusi cekcok lgsg adu bacok ribut utang adu bacok protes pedagang buar pasar pembeli tewas bacok cari makan sahur tewas korban klitih saksikan anaknya tewas keroyok pake sajam acara nikahan acara pembantaian nyawa harganya negara
2	yaallah
3	anak titipan allah jaga dn rawat
...	...
4783	streeesss buibu

3.2.3. Filtering

Tahapan *filtering* digunakan untuk menghapus kata sambung, kata depan dan kata ganti. Pada proses ini dilakukan dengan metode *stopword*. Daftar kata yang digunakan adalah *stoplist* yang berjumlah 758 *stopword* [19]. *Stoplist* yang digunakan dapat dilihat pada Tabel 5.

Tabel 5. *Stoplist*

No	Stopword	No	Stopword	No	Stopword
1	ada	6	akhiri	11	antaranya
2	adalah	7	akhirnya	12	apa
3	adanya	8	aku	13	apaan
4	adapun	9	akulah
5	akhir	10	antara	758	yang

3.2.4. Stemming

Tahapan *stemming* merupakan sebuah proses untuk merubah kata-kata yang berimbuhan menjadi kata dasar. Tahapan ini dilakukan dengan mencari kata dasar pada setiap kata menggunakan modul *sasrawi* pada *phyton*. Hasil dari *stemming* dapat dilihat pada Tabel 6.

Tabel 6. Hasil *Stemming*

No	Komentar / <i>stemming</i>
1	paham indo tumpah darah jadi solusi cekcok lgsg adu bacok ribut utang adu bacok protes dagang buar pasar beli tewas bacok cari makan sahur tewas korban klitih saksi anak tewas keroyok pake sajam acara nikah acara bantai nyawa harga negara
2	yaallah
3	anak titip allah jaga dn rawat
...	...
4783	streeesss buibu

3.3. Pelabelan

Data teks yang didapatkan merupakan data yang tidak berlabel atau *unsupervised* data. Maka dibutuhkan pelabelan untuk merubah data tersebut menjadi *supervised* data. Pada penelitian ini pelabelan data dilakukan oleh seorang ahli bahasa dan pelabelan dilakukan dengan mengklasifikasikan data kedalam tiga kelas yaitu, negatif, netral dan positif [20]. Berikut ini merupakan hasil pelabelan pada dataset dapat dilihat pada Tabel 7.

Tabel 7. Data pelabelan

No	Komentar	Sentimen
1	Gk paham lagi, indo apa ² masalahnya pertumpahan darah di jadikan solusi, cekcok lgsg adu bacok, ribut soal utang adu bacok. Protes pedagang buar di pasar pembeli tewas dibacok, cari makan sahur tewas jadi korban klitih, ibu saksikan anaknya tewas di keroyok pake sajam, acara nikahan jadi acara pembantaian, nyawa orang kek gk ada harganya di negara ini	Netral
2	Ya allah	Positif
3	Anak adalah titipan allah yang harus kita jada dn rawat dgn benar.	Positif
...
4788	Streeess itu buibu	Negatif

3.4. Term Frequency-Invers Document Frequency (TF-IDF)

Dataset yang telah melewati proses pelabelan harus berbentuk angka. *Dataset* tersebut dirubah menjadi angka menggunakan metode TF-IDF. Metode ini bekerja untuk menentukan keterikatan kata (*term*) terhadap dokumen dengan memberikan nilai pada setiap kata [15]. Tujuan metode ini ialah mencari kata-kata yang dapat menunjukkan dokumen teks pada suatu koleksi. TF-IDF menjadi metode terbaik untuk mendapatkan informasi [16] Hasil dari penggunaan TF-IDF dapat dilihat pada Tabel 8.

Tabel 8. Hasil TF-IDF

No	Alas	Allah	Anak	Bantu	Beban	Berat	...	tolong	Tua
1	0.0	0.0	0.32	0.0	0.0	0.0	...	0.0	0.0

2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.81	0.57	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...									
4783	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

3.5. Naive Bayes Classifier (NBC)

NBC merupakan suatu metode klasifikasi data dengan probabilitas sederhana yang menerapkan *teorema bayes* dengan karakter independen yang tinggi. Metode ini sesuai jika digunakan pada dataset yang besar, cepat dalam melakukan klasifikasi dan tingginya akurasi [21]. Pada algoritma NBC hanya dilakukan satu kali percobaan klasifikasi karena NBC tidak memiliki parameter untuk pengujiannya. Sebelum ditambahkan Algoritma NBC, dataset akan dibagi menjadi data latih dan data uji.

3.5.1. Pembagian Data Latih dan Data Uji pada NBC

Pada penelitian ini dilakukan pembagian data latih dan data uji menggunakan *K-Fold Cross Validation* yang telah melewati tahap TF-IDF. Pengujian *K-Fold Cross Validation* pada penelitian ini menggunakan 4783 dataset. Dimana data tersebut akan menjadi data latih dan data uji. Pembagian data latih dan data uji pada penelitian ini digunakan 80:20. Maka didapatkan data latih sebanyak 3826 dan data uji sebanyak 957 data.

3.5.2. K-Fold Cross Validation NBC

Pada tahapan ini dataset akan dipecah sebanyak K bagian, dan akan dilakukan pengujian sebanyak K kali. Hasil percobaan pembagian data latih dan data uji menggunakan *K-Fold Cross Validation* NBC dapat dilihat pada Tabel 9.

Tabel 3. Hasil *K-Fold Cross Validation*

K-Fold	Nilai Akurasi
K-1	80,41%
K-2	85,11%
K-3	85,11%
K-4	89,55%
K-5	78,85%
K-6	83,28%
K-7	84,29%
K-8	91,09%
K-9	87,43%
K-10	84,81%
Nilai Max	91,09%

Berdasarkan Tabel 8, diketahui bahwa percobaan tertinggi terdapat pada K-8 dengan nilai akurasi sebesar 91,09% dan hasil terendah pada K-5 dengan akurasi 78,85%. Hasil *split K-fold Cross Validation* tertinggillah yang akan digunakan pada proses prediksi menggunakan Algoritma NBC yaitu split ke K-8.

3.5.3. Hasil Prediksi Menggunakan NBC

Pada hasil prediksi kelas data uji menggunakan akurasi tertinggi yaitu pada split K-8 dengan nilai akurasi 91,09%. Dikarenakan pada penelitian ini pembagian data latih dan data uji sebesar 80:20 maka, data testing yang digunakan sebanyak 957 data. Hasil prediksi kelas dapat dilihat pada Tabel 10.

Tabel 10. Hasil Prediksi Menggunakan NBC

No	Negatif	Netral	Positif	Hasil prediksi
1.	0.022007	0.019357	0.958636	Positif
2.	0.251154	0.186535	0.562312	Positif
3.	0.022993	0.030276	0.946731	Positif
4.	0.584813	0.109472	0.305715	Negatif
5.	0.060988	0.072704	0.866308	Positif
...
954	0.047641	0.043689	0.908669	Positif
955	0.131208	0.058547	0.810246	Positif
956	0.006755	0.009029	0.984216	Positif
957	0.030060	0.038908	0.931032	Positif

3.6. Prediksi Sentimen

Ini merupakan contoh sub-bab kedua. Isinya dapat disesuaikan dengan kebutuhan.

Setelah dilakukannya prediksi menggunakan algoritma NBC maka didapatkan hasil prediksi sentimen. Yang menggunakan data *testing* sebanyak 957 data dan dikelompokkan ke dalam kelas negatif, netral dan positif. Hasil prediksi sentimen dapat dilihat pada tabel 11.

Tabel 5. Prediksi Sentimen

No	Sentimen	Jumlah	Persentase
1	Negatif	64	6,68%
2	Netral	0	0%
3	Positif	893	93,31%

Pada tabel 10 dapat dilihat bahwa sentimen positif mendominasi sebanyak 93,31% kemudian sentimen negatif didapatkan persentase sebesar 6,68% dan tidak ada sentimen netral. Hasil ini menunjukkan bahwa stigma yang berkembang pada masyarakat mengenai depresi didominasi stigma positif. Hal ini sejalan dengan penelitian yang pernah dilakukan oleh [4] dengan judul "Analisis Sentimen Twitter pada Isu *Mental Health* dengan algoritma Klasifikasi *Naive Bayes*" yang mengatakan bahwa stigma yang berkembang di masyarakat mengenai isu *mental health* didominasi oleh stigma positif.

3.7. Visualisasi Data

Berikut adalah visualisasi data dengan tetap memanfaatkan *python*, visualisasi data ini berupa kumpulan kata berdasarkan *frekuensi* yang dikelompokkan menurut opini negatif, netral dan positif berdasarkan data komentar dari kasus

“seorang ibu tega gorok 3 orang anak nya di brebes”. Hasil dari visualisasi kata berdasarkan opini negatif, netral dan positif dapat dilihat pada Gambar 2 sampai Gambar 4.



Gambar 2. Visualisasi Opini Kata negatif

Berdasarkan Gambar 2 dapat dilihat bahwa kata yang sering muncul pada opini negatif yaitu kata “gila”, “hukum mati”, “anak”, “biadab”, “dasar setan” dan “manusia”.



Gambar 3. Visualisasi Opini Kata Netral

Berdasarkan Gambar 3. Menunjukkan bahwa kata yang sering muncul pada opini netral yaitu kata “gila”, “bunuh”, “anak”, “manusia” dan “iblis”.



Gambar 4. Visualisasi Opini Kata Positif

Berdasarkan Gambar 4 diketahui bahwa kata yang sering ditemukan pada opini Positif yaitu kata “anak” berikutnya disusul oleh kata “suami”, “hidup”, “depresi”, “allah” dan “bunuh anak”..

4. KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah dijelaskan, maka dapat disimpulkan bahwa Hasil pengujian akurasi yang diperoleh pada penelitian ini dengan menggunakan Algoritma NBC adalah 84.11% dengan data *training* sebanyak 3826 sedangkan data *testing* sebanyak 957 data.

Sentimen yang menjadi fokus penelitian yaitu sentimen yang terbagi menjadi positif, netral dan negatif. Dan didapatkan hasil bahwa pada penelitian ini didapatkan hasil sentimen positif mendominasi sebesar 93,31%. Dapat diartikan stigma yang berkembang di masyarakat mengenai penderita depresi didominasi oleh stigma positif.

DAFTAR PUSTAKA

- [1] U. Hasanah, N. L. Fitri, S. Supardi, and L. PH, “Depression Among College Students Due to the COVID-19 Pandemic,” *J. Keperawatan Jiwa*, vol. 8, no. 4, p. 421, 2020.
- [2] B. G. Sudarsono and S. P. Lestari, “Diagnosa Tingkat Depresi Mahasiswa Akhir Terhadap Penelitian Ilmiah Menggunakan Algoritma K-Nearest Neighbor,” *J. Media Inform. Budidarma*, vol. 4, no. 4, pp. 1094–1099, 2020.
- [3] N. Hayatin, “IMPLEMENTASI MULTINOMIAL NAÏVE BAYES UNTUK KLASIFIKASI DATA TWEETS MENGANDUNG TERM,” pp. 344–349, 2020.
- [4] K. Aulia, L. Amelia, and K. Mental, “Analisis Sentimen Twitter Pada Isu Mental Health Dengan Algoritma Klasifikasi Naive Bayes,” vol. 6, no. 2, pp. 60–65, 2020.
- [5] pew research Enter, “No Title,” 2021. [Online]. Available: <https://www.pewresearch.org/topic/internet-technology/platforms-services/social-media/>. [Accessed: 02-Oct-2021].
- [6] M. D. E. Rangkuti, “Analisis topik komentar video beberapa akun youtube e-commerce Indonesia menggunakan metode latent dirichlet allocation,” *Repository.Uinjkt.Ac.Id*, 2020.
- [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.
- [8] E. Yulian, “Text Mining dengan K-Means Clustering pada Tema LGBT dalam Arsip Tweet Masyarakat Kota Bandung,” *J. Mat. “MANTIK,”* vol. 4, no. 1, pp. 53–58, 2018.
- [9] H. Muhamad, C. A. Prasajo, N. A. Sugianto, L. Surtiningsih, and I. Cholissodin,

- “Optimasi Naïve Bayes Classifier Dengan Menggunakan Particle Swarm Optimization Pada Data Iris,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 3, p. 180, 2017.
- [10] G. A. Buntoro, “Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter,” *Integer J.*, vol. 2, no. 1, pp. 32–41, 2017.
- [11] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, “Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm,” *Procedia Comput. Sci.*, vol. 161, pp. 765–772, 2019.
- [12] H. Fan and Y. Qin, “Research on Text Classification Based on Improved TF-IDF Algorithm,” vol. 147, no. Ncce, pp. 501–506, 2018.
- [13] M. A. Rofiqi, A. C. Fauzan, A. P. Agustin, and A. A. Saputra, “Implementasi Term-Frequency Inverse Document Frequency (TF-IDF) Untuk Mencari Relevansi Dokumen Berdasarkan Query,” *Ilk. J. Comput. Sci. Appl. Informatics*, vol. 1, no. 2, pp. 58–64, 2019.
- [14] F. Ratnawati, “Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter,” *INOVTEK Polbeng - Seri Inform.*, vol. 3, no. 1, p. 50, 2018.
- [15] B. Herwijayanti, D. E. Ratnawati, and L. Muflikhah, “Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity,” *Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 1, pp. 306–312, 2018.
- [16] W. Sri, U. Saragih, N. A. Hasibuan, and R. K. Hondroo, “Penerapan Text Mining Dengan Menggunakan Metode TF-IDF Untuk Menentukan Genre Dari Komik,” vol. 4, pp. 191–199, 2020.
- [17] A. Muzaki and A. Witanti, “Sentiment Analysis of the Community in the Twitter To the 2020 Election in Pandemic Covid-19 By Method Naive Bayes Classifier,” *J. Tek. Inform.*, vol. 2, no. 2, pp. 101–107, 2021.
- [18] M. R. Fadhillah, I. Ishak, and P. S. Ramadhan, “Implementasi Sistem Pakar Mendiagnosa Penyakit Penyakit Gastritis Dengan Menggunakan Metode Teorema Bayes,” *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 4, no. 1, p. 1, 2021.
- [19] F. Z. Tala, “A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia,” *M.Sc. Thesis, Append. D*, vol. pp, pp. 39–46, 2003.
- [20] P. O. A. Sunarya, R. Refianti, A. B. Mutiara, and W. Octaviani, “Comparison of accuracy between convolutional neural networks and Naïve Bayes Classifiers in sentiment analysis on Twitter,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 5, pp. 77–86, 2019.
- [21] D. Normawati and S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, 2021.