

IMPLEMENTASI DATA MINING DALAM KASUS MENTAL HEALTH PADA SOSIAL MEDIA TWITTER MENGGUNAKAN METODE NAIVE BAYES

IMPLEMENTATION OF DATA MINING IN THE CASE OF MENTAL HEALTH ON SOCIAL MEDIA TWITTER USING NAIVE BAYES METHOD

Yunis Femilia Nugraini¹, Rd. Rohmat Saedudin², Rachmadita Andreswari³

^{1,2,3} Universitas Telkom, Bandung

yunisnis@student.telkomuniversity.ac.id¹, rdrohmat@telkomuniversity.ac.id²,

andreswari@telkomuniversity.ac.id³

Abstrak

Kesehatan mental, khususnya di Indonesia kesehatan mental masih jauh dari perhatian pemerintah maupun masyarakat, masih minimnya kesadaran masyarakat bahwa kesehatan itu bukan hanya tentang fisik saja tapi juga perlu memperhatikan kesehatan mental. Masyarakat Indonesia masih sangat tabu dalam menyikapi hal ini, peningkatan jumlah masyarakat yang mengalami permasalahan kesehatan mental selalu mengalami kenaikan dari tahun ke tahun, masyarakat masih menganggap bahwa kesehatan mental tidak terlalu berbahaya sehingga masih di anggap remeh dalam menangani kasus yang seperti ini, padahal setiap tahunnya pasti ada saja kasus bunuh diri dikarenakan depresi. Dari penelitian ini akan mengetahui seberapa pentingnya kesehatan mental yang dianalisis dari pengguna Twitter, dengan memberikan pelabelan positif, negatif dan juga netral. Pada penelitian ini algoritma yang digunakan adalah Algoritma *Naive Bayes*, menggunakan algoritma ini karena memiliki nilai akurasi yang cukup baik dan juga simple dalam proses pengklasifikasian. Penelitian ini menggunakan rasio 70:30 dengan nilai akurasi 89%. Pada proses penelitian ini melakukan tahap crawling data, preprocessing, pembobotan kata, klasifikasi data, dan kemudian performansi evaluasi. Hasil penelitian menunjukkan bahwa algoritma *Naive Bayes* memiliki nilai *macro average* untuk *precision*, *recall*, dan *f1-score* dengan nilai 63% hingga 74%, dan *weighted average* untuk *precision*, *recall*, dan *f1-score* dengan nilai 89% hingga 92%.

Kata Kunci : *Naive Bayes, Data Mining, Kesehatan Mental, Twitter*

Abstract

Mental health, especially in Indonesia, mental health is still far from the attention of the government and the community, there is still a lack of public awareness that health is not only about physical but also needs to pay attention to mental health. Indonesian people are still very taboo in responding to this, the increasing number of people who experience mental health problems always increases from year to year, people still think that mental health is not too dangerous so it is still underestimated in handling cases like this, even though every year there must be cases of suicide due to depression. From this research, we will find out how important mental health is analyzed from Twitter users, by giving positive, negative and neutral labeling. In this study, the algorithm used is the Naive Bayes Algorithm, using this algorithm because it has a fairly good accuracy value and is also simple in the classification process. This study uses a ratio of 70:30 with an accuracy value of 89%. In this research process, data crawling, preprocessing, word weighting, data classification, and performance evaluation stages are carried out. The results show that the Naive Bayes algorithm has a macro average value for precision, recall, and f1-score with a value of 63% to 74%, and a weighted average for precision, recall, and f1-score with a value of 89% to 92%.

Keywords: *Naive Bayes, Data Mining, Mental Health, Twitter*

1. Pendahuluan

Perkembangan teknologi internet yang begitu pesat ini, kini banyak melahirkan berbagai jenis media sosial. Hal ini didukung pula dengan perkembangan teknologi telepon selular yang menyediakan fasilitas bagi pengguna media sosial ini. Media sosial ini menjadi sangat populer karena memberikan kemudahan-kemudahan dalam melakukan komunikasi. Media sosial yang paling populer digunakan baik anak-anak maupun dewasa antara lain, Facebook, Twitter, Instagram, WhatsApp, You Tube dan masih banyak lagi (Warpindyastuti & Sulistyawati, 2018).

Salah satu media sosial yang banyak digunakan masyarakat adalah Twitter. Berdasarkan buku yang ditulis oleh Hadi (2010: 2)[13] pengertian Twitter adalah situs microblog yang memberikan fasilitas bagi pengguna untuk mengirimkan sebuah pesan teks dengan panjang maksimal 140 karakter melalui SMS, pengirim pesan instan, surat elektronik.

Media social twitter juga membuat fitur pengguna untuk dapat memberikan pendapatnya dengan cara mempublikasikan tweet mereka atau sekedar memberikan komentar, dengan demikian mereka dapat membahas apapun mengenai informasi maupun isu yang sedang di perbincangkan.

Berdasarkan dari data-data komentar dan juga tweet pengguna twitter berdasarkan dari pendapat masyarakat yang menggunakan twitter, dapat dilakukan analisis sentiment guna untuk menyaring tweet ataupun komentar yang terkait dengan isu kesehatan mental, yang nantinya akan dikelompokkan menjadi komentar negative, positif, dan juga netral.

Dalam melakukan analisis sentiment, peneliti menggunakan algoritma *naïve bayes* dalam proses mengklasifikasi data twitter mengenai kesehatan mental.

2. Metode Penelitian

2.1 Model Konseptual

Penelitian yang penulis lakukan ini akan menghasilkan sebuah informasi mengenai kesehatan mental atau *Mental Health* di Indonesia, dengan menentukan atribut yang dibutuhkan dalam melakukan proses pengelolaan data ini. Pada proses penelitian ini penulis melakukan identifikasi masalah terlebih dahulu, kemudian membuat rumusan masalah yang menjadi titik sentral dari suatu penelitian, kemudian masuk ke pengambilan data, penulis mengambil data dari *Twitter*, penulis mendapatkan data dari *Twitter* sesuai dengan *keyword* yang penulis inginkan, setelah masuk ke tahap analisis data untuk mendeskripsikan suatu data agar lebih mudah di pahami, kemudian masuk ke proses pengelolaan data, dan setelah data sudah berhasil untuk di kelola kemudian masuk ke kesimpulan dan saran.

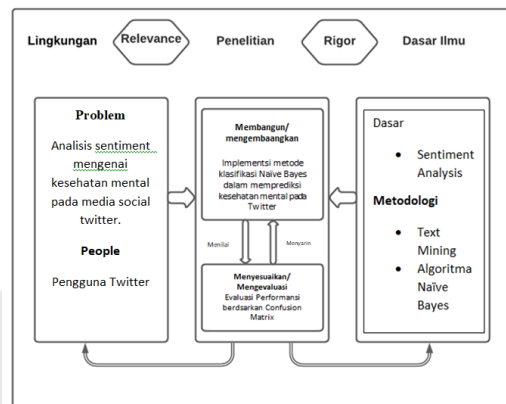


Figure 2-1 Model Konseptual

Pada Gambar III-1, dijelaskan pada lingkungan aspek people yaitu pengguna media social Twitter, pengguna Twitter yang dimaksudkan adalah yang memberikan komentar atau tweet yang berkaitan dengan kesehatan mental (*Mental Health*). Kemudian problemnya adalah analisis sentiment mengenai kesehatan mental pada media social twitter.

2.2 Sistematisa Penyelesaian Masalah

Penelitian ini menggunakan sistematisa penelitian dengan beberapa tahapan, yaitu inisialisasi, implementasi dan hasil dan kesimpulan.

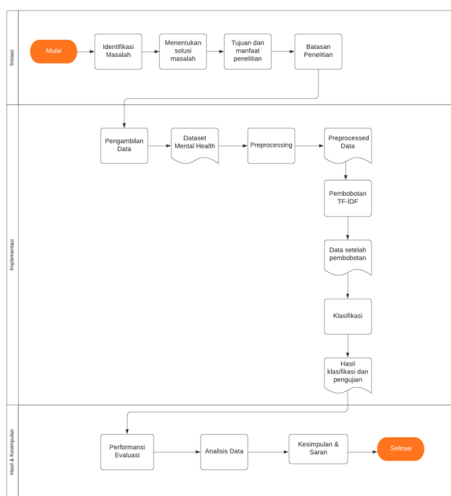


Figure 2-II Sistematisa Penyelesaian

2.2.1 Inisiasi

Pada tahap ini melakukan identifikasi masalah terkait dengan topik yang diambil agar sesuai dengan proses dan hasil akhir, dengan menentukan solusi dari masalah yang diambil terkait dengan identifikasi masalah.

2.2.2 Implementasi

Pada tahap implementasi, peneliti melakukan crawling data twitter terlebih dahulu. Crawling data untuk mendapatkan data yang sesuai dengan topik yang nantinya akan di proses, crawling data menggunakan kata kunci *#MentalHealth*. Untuk penelitian ini komentar atau tweet yang diambil adalah 1058 komentar.

2.2.3 Hasil & Kesimpulan

Pada tahap ini dilakukan proses performansi evaluasi dengan melakukan pemetaan hasil prediksi ke table *confusion matrix*, dengan adanya

confusion matrix mempermudah untuk melakukan perhitungan *precision*, *recall*, *f1-score*, dan *accuracy*.

Kemudian diakhiri dengan menghitung akurasi dari model yang telah dibuat. Tahapan evaluasi berakhir dengan menganalisis hasil dari performansi evaluasi, dan dari hasil itu dapat dilakukan penarikan kesimpulan.

3. Hasil dan Pembahasan

3.1 Dataset

Dalam pengujian ini penulis menggunakan *tweet-tweet* yang sebelumnya sudah didapatkan dari twitter pada halaman resmi Twitter yang didapatkan dengan cara *crawling* menggunakan jupyternotebook, peneliti menggunakan bahasa pemrograman python dalam mengelola data. Data yang terkumpul sebanyak 1058 data, setelah itu data akan dilakukan proses preprocessing, dan dilakukan proses pelabelan komentar menggunakan Textblob.

Table 3-1 Pelabelan Komentar

No	Komentar	Label
1	Malemnya nangis paginya senyum kayak tadi malem its okay aja gitu, gaada yg harus dikhawatirin, hal yg wajar banget.	Negatif
2	Perlu disadari bahwa mereka yang kehilangan seseorang akibat bunuh diri memiliki risiko lebih tinggi	Positif
3	Susah banget ngendaliin diri sendiri biar gak mikirin sesuatu berlebihan gak ngerasa cemas berlebihan gak takut	Netral

Table 3-1 diatas merupakan dataset yang sudah dilakukan pelabelan terdiri dari positif, negative dan netral, dari 1058 dataset terdapat 88 yang bernilai positif, 25 bernilai negative, dan 945 bernilai netral, dari hasil tersebut artinya bahwa tweet dari data twitter mengenai kesehatan mental itu bernilai netral.

3.2 Hasil Klasifikasi

Pada proses klasifikasi *naïve bayes* ini, menggunakan rasio 70:30, data terbagi menjadi 2 bagian yaitu data training dan juga data testing. Data testing berfungsi untuk menguji tingkat akurasi suatu

data dari model yang dibuat data testing, sedangkan data training dapat mempengaruhi suatu tingkat akurasi data.

Table 3-2 Ratio

Rasio	Hasil Akurasi	Data Training	Data Testing	Total Data
70 : 30	89%	740	318	1058

Dengan model *classifier naïve bayes* memiliki performa akurasi dengan nilai 89%, dengan data training sebanyak 740 tweet dan untuk data testing sebanyak 318 tweet.

3.3 Analisis Sentimen

Pada proses ini melakukan analisis sentiment menggunakan data testing sebanyak 318 tweet, pada hasil prediksi metode naïve bayes dengan memberikan label positif, negative, dan netral sesuai dengan jumlah data testing. Dari hasil klasifikasi yang sudah dilakukan menghasilkan klasifikasi sentiment seperti gambar :

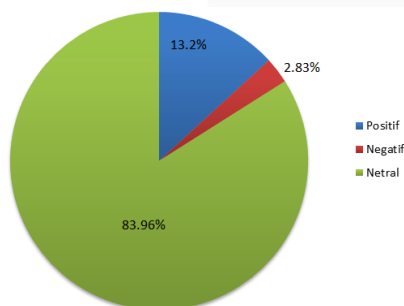


Figure 3-I Grafik Pie Chart Hasil Klasifikasi

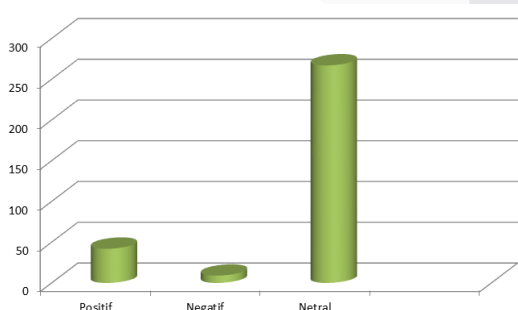


Figure 3-II Grafik Bar Chart Hasil Klasifikasi

Pada gambar grafik hasil klasifikasi pada data testing terhadap tweet twitter mengenai “Mental Health” mencapai

267 prediksi label positif, 9 prediksi label negative, dan 42 prediksi label positif. Dan pada gambar V-1 prediksi label netral mendominasi dengan total 89.62% dari total data *testing*. Dari hasil tersebut dapat disimpulkan bahwa pengguna twitter banyak memberikan tweet netral.

3.4 Evaluasi Performansi

Setelah berhasil melakukan proses prediksi pada data testing, selanjutnya akan melakukan evaluasi performansi untuk mengukur kinerja model *classifier naïve bayes* yang telah di buat, dari hasil prediksi yang sebelumnya perlu adanya *confusion matrix*, peneliti menggunakan *package matrices* untuk mempermudah proses perhitungan *confusion matrix* dan performansi evaluasi. Berikut merupakan *confusion matrix* dari *classifier naïve bayes* :

Table 3-3 Pengujian Confusion Matrix

Actual \ Predicted	Positive	Negative	Neutral	Total
Positive	20	2	20	42
Negative	0	4	5	9
Neutral	5	2	260	267
Total	25	8	285	318

Berdasarkan Tabel 3-3, model *classifier naïve bayes* memprediksi label positif dengan benar (*True Positive*) sebanyak 20, label positif yang salah (*False Positive*) sebanyak 2, dan label positif yang netral (*Neutral Positive*) sebanyak 20, dengan total 42 total prediksi label positif sesuai dengan grafik di gambar 1. Untuk hasil prediksi label negative dengan benar (*True Negative*) sebanyak 0, prediksi label negative yang salah (*False Negative*) sebanyak 4, dan prediksi label negative yang netral (*Neutral Negative*) sebanyak 5, dengan total 9 total prediksi label negative sesuai dengan grafik di gambar. Sedangkan untuk hasil prediksi label netral dengan benar (*True Neutral*) sebanyak 5, prediksi label netral yang salah (*False Neutral*)

sebanyak 2, dan prediksi label netral yang netral (*Neutral Neutral*) sebanyak 260, dengan total 267 total prediksi label netral sesuai dengan grafik di gambar 1. Ketiga label positif, negative, dan netral hasil prediksi sebelumnya telah dipetakan pada *confusion matrix*, selanjutnya peneliti melakukan evaluasi performansi dari *confusion matrix* yang telah dibuat. Pada Tabel V-4 adalah hasil dari evaluasi performansi dengan total 318 jumlah data testing :

Table 3-4 Hasil Evaluasi Performansi

Actual \ Predicted	Precision	Recall	F1-Score
Positive Label	48%	80%	60%
Negative Label	44%	50%	47%
Neutral Label	97%	91%	94%
Macro Average	63%	74%	67%
Weighted Average	92%	89%	90%
Accuracy	89%		

Pada Tabel 3-4 terdapat 3 label yaitu positif, negative dan netral beserta dengan nilai evaluasi performansinya, dari hasil tersebut dapat dilihat bahwa untuk label positif, nilai precision yang diperoleh mencapai 48%, nilai recall sebesar 80%, dan F1-Score sebesar 60%. Untuk label negative, nilai precision yang diperoleh yaitu 44%, nilai recall sebesar 50%, dan F1-Score 47%. Sedangkan untuk label netral, nilai precision yang diperoleh mencapai 97%, nilai recall sebesar 91%, dan untuk nilai F1-Score sebesar 94%. Hasil evaluasi performansi dengan rasio data testing 70:30 mencapai akurasi 89%.

Hasil evaluasi performansi dengan classifier Naïve Bayes dirata-rata kan dengan Macro Average dan Weighted Average, perbedaan kedua rata-rata tersebut yaitu Macro Average menggunakan pembagi dengan jumlah label yang diklasifikasi.

3.5 Wordcloud

Pada penelitian ini, peneliti membuat visualisasi dalam bentuk wordcloud, wordcloud akan menampilkan kata-kata yang sering digunakan oleh pengguna twitter dalam memberikan komentarnya

mengenai tentang kesehatan mental, sehingga mempermudah dalam melihat informasi mengenai kata apa saja yang sering muncul. Hasil wordcloud dapat dilihat pada Gambar 3-III :

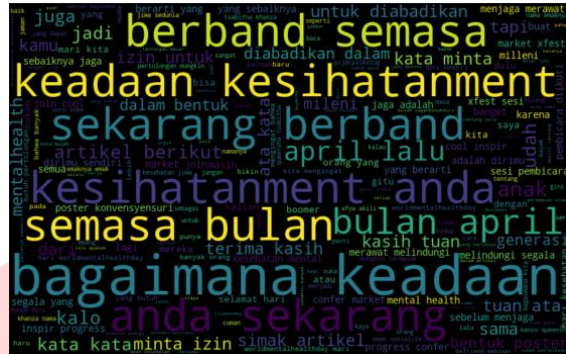


Figure 3-III Wordcloud

Dari hasil Gambar V-3 wordcloud yang bisa dilihat, ada beberapakata yang sering digunakan pengguna twitter dalam mengomentari topic terkait kesehatan mental, seperti terdapat kata “Keadaan”, “Kesehatanmental”, dll. Wordcloud ini adalah gambaran secara umum mengenai kata yang paling banyak digunakan oleh pengguna twitter dalam memberi komentar mengenai kesehatan mental di Twitter.

4. Kesimpulan

1. Analisis sentiment yang dilakukan dengan menggunakan data tweet dari pengguna Twitter terhadap kesehatan mental dengan total 1058 tweet. Dengan menggunakan rasio *training* dan *testing* 70:30, maka klasifikasi dengan algoritma Naïve Bayes hanya akan memprediksi data dengan jumlah 318 tweet. Hasil prediksi yang sudah peneliti lakukan memperoleh label prediksi positif dengan total 42, label prediksi negative dengan total 9, dan label prediksi netral dengan total 267. Dari Tabel 2 dapat dilihat bahwa *clasiffier* menebak sejumlah 20 label positif secara benar (True Positive), 4 menebak negative secara benar (True Negative), dan 260 menebak netral secara benar (True Neutral). Dengan hasil tersebut membuktikan bahwa model yang dibuat

dengan rasio 70:30 secara mayoritas dapat melakukan prediksi dengan benar, maka dapat di simpulkan analisis sentiment menggunakan algoritma *Naïve bayes* menghasilkan sentiment positif sebanyak 7.86%, negative sebanyak 2.51%, dan netral sebanyak 89.62%. Maka topik mengenai kesehatan mental di Twitter bernilai netral.

2. Hasil performansi evaluasi dari klasifikasi menggunakan algoritma *Naïve Bayes* menggunakan rasio 70:30 sebagai acuan performa untuk melakukan prediksi. Pemetaan hasil prediksi terhadap label aktual menggunakan *confusion matrix*. Hasil evaluasi performansi terdapat nilai *recall*, *precision* dan *f1-score* untuk label positif 40% hingga 80%. Untuk label negative, terdapat nilai *recall*, *precision*, dan *f1-score* dengan nilai 44% hingga 50%. Sedangkan untuk label netral, terdapat nilai *recall*, *precision*, dan *f1-score* dengan nilai 91 hingga 97 %. Peneliti juga melakukan *macro average* dan *weighted average* pada hasil evaluasi performansi. Didapat *macro average* dengan rentang nilai 63% hingga 74%, dan *weighted average* dengan nilai 89% hingga 92%. Maka dari hasil tersebut peneliti menarik kesimpulan bahwa performa algoritma *Naïve Bayes* sudah cukup baik. Hal ini dibuktikan dengan nilai akurasi yang cukup baik yaitu 89%.
3. Hasil dari performansi dapat diartikan bahwa dampak isu kesehatan mental terhadap komentar dari twitter adalah netral. Setiap label memiliki nilai yang berbeda dimana positif adalah seseorang yang setuju akan pentingnya kesehatan mental dan penanganannya, negative adalah seseorang yang tidak setuju akan pentingnya kesehatan mental, dan netral adalah seseorang yang tidak paham mengenai kesehatan mental dan masih bingung harus bertindak seperti apa.

Referensi

- [1] Warpindyastuti, L.D., & Sulistyawati, M.E., (2018). Pemanfaatan Teknologi Internet Menggunakan Media Sosial Sebagai Sarana Penyebaran Informasi dan Promosi Pada MIN 18 Jakarta. Widya Cipta (Vol II No. 1).
- [2] Malfasari, E., Sarimah, Febtriana, R., Herniyanti, R., (2020), KONDISI MENTAL EMOSONAL PADA REMAJA. Jurnal Keperawatan Jiwa, (Volume 8 No 3, Agustus 2020, pp. 241 – 246).
- [3] Simorangkir, H., & Lhaksmana, M.K., 2018. Analisis Sentimen Pada Twitter Untuk Games Online Mobile Legends Dan Arena Of Valor Dengan Metode Naïve Bayes Classifier. eProceedings of Engineering (vol. 5, No. 3).
- [4] Andika, A.L., Azizah, N.P., & Respatiwan, (2019) Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier. Indonesian Journal of Applied Statistics (Volume 2 No. 1 May 2019).
- [5] Migunani & Aditama, K., (2020), PEMANFAATAN NATURAL LANGUAGE PROCESSING DAN PATTERN MATCHING DALAM PEMBELAJARAN MELALUI GURU VIRTUAL. ELKOM(vol. 13, No. 1, pp. 121-133).