

ANALISIS SENTIMEN MASYARAKAT TERHADAP UNIVERSITAS BUANA PERJUANGAN KARAWANG DENGAN ALGORITME SVM dan NAÏVE BAYES

¹Amril Mutoi Siregar

²Sutan Faisal

³Banani Widiharto

Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Buana Perjuangan
Karawang, Jl Ronggowaluyo, Karawang 41311, Indonesia

amrilmutoi@ubpkarawang.ac.id, Sutanfaisal@ubpkarawang.ac.id,
if17.bananiwidiharto@mhs.ubpkarawang.ac.id.

ABSTRAK

Salah satu tugas dunia kampus adalah untuk menghasilkan lulusan yang siap kerja digunakan oleh industry dan menciptakan pengusaha baru, tentunya Universitas Buana Perjuangan Karawang juga harus meningkatkan lulusan secara menyeluruh, termasuk juga dalam upaya mengetahui respon atau mendapat masukan dari masyarakat terkait dengan kehadiran UBPK. Sehingga dalam penelitian ini dirumuskan bagaimana membuat peta analisis sentiment yang berisikan respon/feedback/komentar masyarakat dalam media sosial terkait dengan berbagai hal yang tentang Universitas Buana Perjuangan Karawang. Pendekatan yang dilakukan dalam penyelesaian masalah dimulai dari pengumpulan data, preprocessing (cleansing, tokenizing, dan case folding), dan klasifikasi. hasil yang diharapkan penelitian ini ada sebagai input terhadap kampus untuk menjadi eksistensi khusus di kabupaten karawang. pengumpulan data yang terkumpulkan sebanyak 309 twitt dapat dikategorikan bahwa perbincangan di media sosial twitter tidak populer. Sehingga model yang dibangun dimaksimal seperti dengan algoritme SVM dengan akurasi training dan testing 86.66% dan 76.92% serta algoritme naïve bayes dengan akurasi 96.66% dan 69.23%. Dapat diambil kesimpulan bahwa hasil yang didapatkan oleh model kurang baik.

Kata kunci:, Analisis Sentimen, Twitter, Text Mining, UBPK, SVM, Naïve Bayes

ABSTRACT

One of the tasks of the campus world is to produce graduates who are ready to work to be used by industry and create new entrepreneurs, of course, Universitas Buana Perjuangan Karawang must also improve graduates as a whole, including in an effort to find out the response or get input from the public regarding the presence of UBPK. So that in this study it is formulated how to make a sentiment analysis map containing community responses/feedback/comments on social media related to various things about Universitas Buana Perjuangan Karawang. The approach taken in solving the problem starts from data collection, preprocessing (cleansing, tokenizing, and case folding), and classification. the expected results of this study are as input to the campus to become a special presence in the Karawang district. data collection collected as many as 309 twitts can be categorized that the conversation on twitter social media is not

Karawang, 28 Februari 2023

popular. So that the model built is maximized as with the SVM algorithm with 86.66% and 76.92% training and testing accuracy and the naïve bayes algorithm with 96.66% and 69.23% accuracy. It can be concluded that the results obtained by the model are not good enough.

Keywords: *Sentiment Analysis, Twitter, Text Mining, UBPK, SVM, Naïve Bayes*

PENDAHULUAN

Siklus yang terjadi dalam Sistem Penjaminan Mutu Internal (SPMI) adalah Penetapan, Pelaksanaan, Evaluasi, Pengendalian, dan Peningkatan (PPEPP). Dalam menjalankan tugasnya sebagai pencetak lulusan yang siap kerja dan digunakan oleh industri, tentunya UMBY juga harus menerapkannya secara menyeluruh, termasuk juga dalam upaya mengetahui respon masyarakat terhadap berbagai hal. Salah satu hal yang dapat dilakukan adalah dengan mengetahui respon atau mendapat masukan dari masyarakat terkait dengan eksistensi UMBY. Mulai era 4.0 masyarakat cenderung menghabiskan waktunya dengan sosial media, twitter misalnya. Pada aplikasi tersebut masyarakat bisa saja memberikan tweet dengan keyword tertentu yang menyangkut dengan UMBY. Teknik penarikan beberapa komentar tersebut dikenal dengan Text Mining yang difokuskan ke dalam Sentiment Analysis. Harapannya setelah ditemukannya peta analisis sentimen ini maka pihak UMBY dapat melakukan evaluasi dan tindak lanjut yang diperlukan.

Berdasarkan latar belakang tersebut, maka rumusan permasalahan dalam penelitian ini adalah bagaimana membuat peta analisis sentimen yang berisikan respon/feedback/ komentar masyarakat dalam media sosial terkait dengan berbagai hal yang ada di UMBY, yang bertujuan untuk merancang peta analisis sentimen opini publik terhadap UMBY. Selanjutnya berdasarkan peta analisis tersebut akan dilakukan tindak lanjut yang dalam tahap ini masuk dalam langkah Evaluasi (PPEPP). Selanjutnya hasil dari penelitian ini diharapkan dapat digunakan oleh pihak UMBY sebagai masukan dan dapat melakukan tindak lanjut yang sesuai. Berdasarkan penelitian sebelumnya mengenai sentiment analisis yang diterapkan oleh kampus sebagai bahan masukan untuk meningkatkan layanan. Penelitian ini menganalisa sentimen analisis pendapat masyarakat terhadap kehadiran kampus di karawang, dengan menggunakan pendekatan teks mining dengan algoritme support vector machine (SVM) dan naïve bayes yang diukur dengan beberapa model evaluasi akurasi, confusion matriks dan ROC, agar mudah dipahami hasil dari penelitian ini.

TINJAUAN PUSTAKA

Beberapa penelitian yang terkait dalam penelitian ini antara lain seperti penelitian mengenai analisis sentimen twitter untuk teks berbahasa indonesia dalam penelitian ini dilakukan proses klasifikasi dokumen tekstual ke dalam dua kelas, yaitu kelas sentimen positif dan negatif. Data yang digunakan berasal dari jejaring sosial Twitter berdasarkan query dalam Bahasa Indonesia. Penelitian ini bertujuan untuk menentukan sentimen publik terhadap objek tertentu yang disampaikan di Twitter dalam bahasa Indonesia, sehingga membantu usaha untuk melakukan riset pasar atas opini publik.

Dalam penelitian ini digunakan Algoritma Maximum Entropy untuk POS tagger dan Support Vector Machine untuk membangun model klasifikasi atas data pelatihan. Hasil dari implementasi klasifikasi diperoleh akurasi 86,81 % pada pengujian 7 fold cross validation untuk tipe kernel Sigmoid. Pelabelan kelas secara manual dengan POS tagger menghasilkan akurasi 81,67% [1]. Selanjutnya penelitian mengenai twitter sentimen analysis terhadap brand

reputation. Dalam penelitian ini dilakukan perhitungan reputasi produk XL Axiata, dan dibandingkan dengan produk Telkomsel dan Indosat. Metode yang dilakukan dengan ekstrak data, analisis sentimen. Hasil dalam penelitian ini menunjukkan bahwa rata-rata perbandingan produk, XL Axiata memperoleh rata-rata nilai reputasi melebihi dibandingkan dengan Telkomsel dan Indosat yaitu sebesar 24.5%, sedangkan Telkomsel 13.2% dan Indosat 19.3% [2]. Selanjutnya penelitian mengenai Sentimen Twitter di SAMSAT. Penelitian ini bertujuan untuk membangun aplikasi analisis sentiment yang menerapkan pendekatan Naïve Bayes Classifier untuk mengklasifikasikan kata-kata dan difokuskan pada tweet dalam Bahasa Indonesia. Data diperoleh melalui cara web scrapping dan sumber teks yang digunakan sebagai topik bahasan adalah Sistem Administrasi Manunggal Satu Atap (SAMSAT) Malang Kota. Proses klasifikasi dilakukan melalui serangkaian tahapan seperti preproses (case folding, cleaning, tokenizing, dan stopword) serta proses klasifikasi dengan algoritma Naïve Bayes Classifier itu sendiri untuk mendapatkan hasil klasifikasi dengan kategori positif, negatif atau netral. Hasil dari penelitian ini menunjukkan bahwa algoritma Naïve Bayes Classifier memberikan unjuk kerja yang baik dalam analisis sentimen. Dari hasil uji akurasi klasifikasi yang dilakukan oleh aplikasi menghasilkan nilai akurasi tertinggi pada setiap kategori positif, negatif, netral masing-masing sebesar 82%, 92%, 80% dengan jumlah data latih 200 tweet negatif, 200 tweet positif, dan 200 tweet netral [3].

Selanjutnya penelitian mengenai analisis sentimen terhadap pemerintahan Joko Widodo. Dalam penelitian ini analisis sentimen dilakukan dengan proses klasifikasi dokumen tekstual ke dalam dua kelas, yaitu kelas sentimen negatif dan positif. Data opini diperoleh dari jejaring social Twitter berupa tweet berdasarkan query dalam Bahasa Indonesia. Data yang digunakan berjumlah 400 tweet terdiri dari 300 data latih dan 100 data uji. Data latih merupakan data yang telah diketahui sentimennya, 300 data latih terdiri dari 150 data kelas sentimen negatif dan 150 data kelas sentimen positif. Penelitian ini bertujuan untuk menentukan tweet merupakan tweet positif atau negatif yang disampaikan di Twitter dalam Bahasa Indonesia. Pengklasifikasian data tweet menggunakan algoritma naïve bayes classifier. Hasil klasifikasi pada data uji menunjukkan, algoritma Naïve Bayes Classifier memberikan nilai akurasi sebesar 97%. Untuk nilai akurasi tiap sentimennya yaitu 96% untuk sentiment negatif dan 98% untuk sentimen positif [4].

Selanjutnya penelitian mengenai sentiment analysis terhadap tweet bernada sarkasme berbahasa Indonesia. pada penelitian ini dilakukan dibahas mengenai sentimen analisis bernada sarkasme pada Tweet berbahasa Indonesia dengan menggunakan fitur interjeksi dan unigram sebagai fitur utama mendeteksi kalimat sarkasme serta membandingkan 2 metode klasifikasi yaitu Naive Bayes dan Support Vector Machine dengan kernel polinomial. Hasil dari penelitian ini menunjukkan penggunaan fitur interjeksi dan unigram sebagai pendeteksian sarkasme pada tweet berbahasa Indonesia mampu meningkatkan akurasi dengan rata-rata kenaikan akurasi lebih dari 8% untuk classifier Naive Bayes dan lebih dari 13% untuk classifier Support Vector Machine dibandingkan hanya menggunakan fitur unigram saja. Hasil lainnya adalah akurasi terbaik adalah metode Naive Bayes dengan akurasi terbaik yang diperoleh mencapai lebih dari 91 [5]. Selanjutnya penelitian mengenai sentiment analisis twitter terhadap suatu kebijakan pemerintah. Dalam penelitian ini dilakukan proses data mining (transform, tokenize, stemming, dan classification). Penelitian ini bertujuan untuk mengklasifikasikan sentiment positif, netral dan negatif dari twitter terhadap dua kata kunci yaitu “Jokowi” dan “Covid”. Dalam penelitian digunakan metode Naive Bayes, Support Vector Machine, dan k-NN untuk klasifikasi, dengan nilai terbaik yang dihasilkan yaitu Support Vector Machine dengan accuracy 84.58%, precision 82.14% dan recall 85.82% [6].

METODE PENELITIAN

Jenis Penelitian

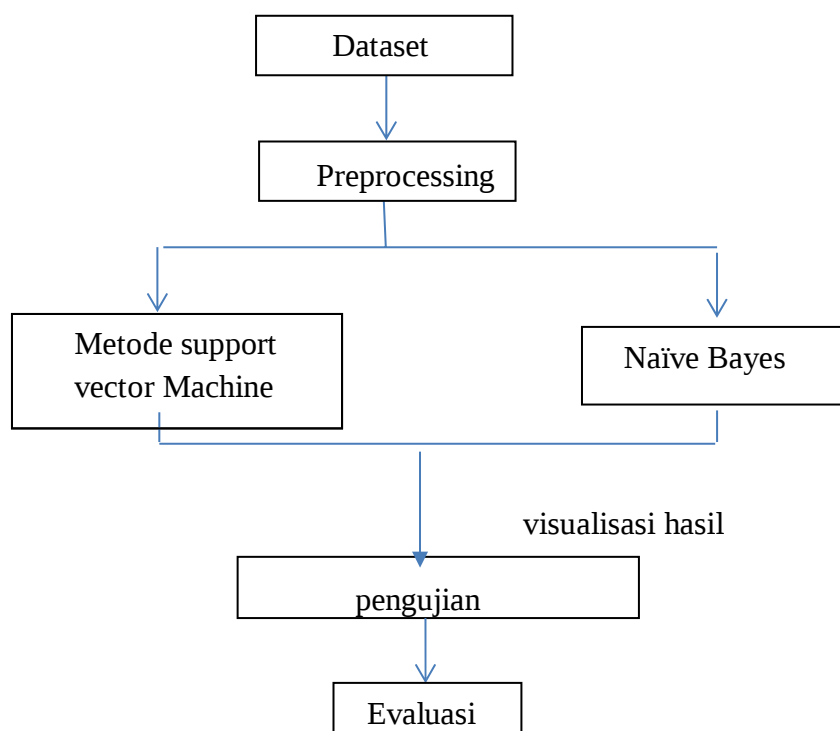
Penelitian ini termasuk jenis kuantitatif merupakan penelitian ilmiah yang sistematis terhadap bagian-bagian dan fenomena serta kausalitas hubungan-hubungannya antara parameter dan penelitian ini juga sebagian besar dilakukan dengan menggunakan metode statistik yang digunakan untuk mengumpulkan data kuantitatif dari studi penelitian. *Sentimen analysis* adalah salah satu cabang penelitian *text mining*. Analisis sentiment berkaitan dengan bidang yang lebih luas seperti pengolahan bahasa alami, komputasi linguistik, dan *text mining* untuk tujuan menganalisis, sentiment, pendapat, sikap, evaluasi, penilaian dan emosi seseorang berkenaan pada topik, produk, layanan, organisasi, atau kegiatan tertentu [7]. Tugas dasar dalam analisis sentimen adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, apakah pendapat yang dikemukakan dalam dokumen bersifat positif, negatif atau netral

Waktu dan Tempat Penelitian

Pengambilan data pada penelitian ini dilakukan pada bulan februari 2023 dengan data yang diambil pada website twitter pada rentang waktu januari 2020 - 2022 dengan banyak sentiment yang dapat terdownload sebanyak 309 dengan 5 atribut. Target penelitian ini untuk mendapatkan model klasifikasi yang paling optimal [8].

Teknik Analisis Data

Metode analisa data yang digunakan dalam penelitian ini adalah dengan membuat alur penelitian dengan jelas dan sistematis. Langkah yang dilakukan pada penelitian sebelumnya adalah dapat digambarkan pada kerangka pikir penelitian yang dapat dilihat pada gambar 3.



Gambar 1. Metode yang digunakan

Penjelasan dari diagram alir diatas adalah sebagai berikut:

Dataset

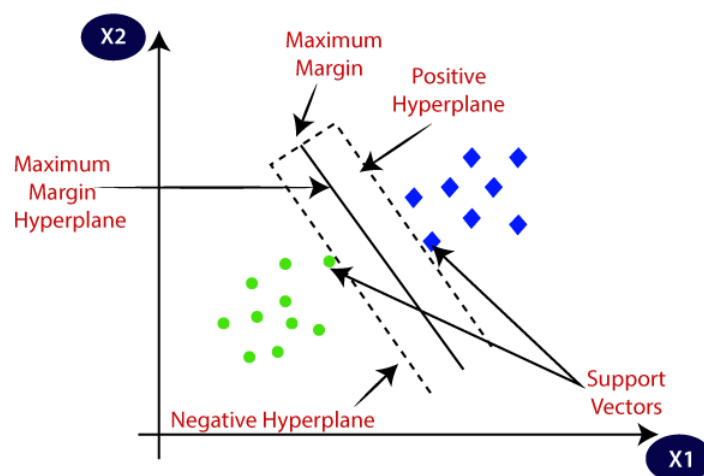
Data diinput sebagai dataset pada penelitian ini adalah data yang dikumpulkan dari twitter.com dengan hashtag #UBPK # buana perjuangan karawang #layanan UBP karawang, dan lain lain.

Pre-processing data

Tahap *preprocessing*, *tweet* yang telah dikumpulkan akan melalui beberapa tahap pemrosesan teks yang terdiri dari *cleansing*, *tokenizing*, *case folding*. Tujuan dari pemrosesan teks adalah supaya data yang didapat akan lebih terstruktur agar lebih mudah untuk dilakukan pengolahan data. Cleaning Proses cleaning yang dilakukan yaitu merubah teks menjadi huruf kecil (*case folding*), menghapus karakter selain huruf, menghapus username pengguna atau mentions (@), menghapus hashtag (#), menghapus URL atau link dari setiap komentar. Filtering Pembuangan kata-kata tidak penting dari hasil token. Selain juga dilakukan penghapusan tanda baca dan stopwords. Stopword diproses pada sebuah kalimat jika mengandung kata-kata yang sering keluar dan di anggap tidak penting seperti waktu, penghubung, dan lain sebagainya. Stemming Pada tahapan stemming, yaitu pengubahan kata berimbuhan menjadi kata dasar. Tokenizing Pada tahapan tokenizing, setiap kata akan dipisahkan berdasarkan spasi yang ditemukan [9].

Metode Support vector Machine

Untuk mendukung Vector Machine atau SVM adalah salah satu algoritma Pembelajaran Terawasi yang paling populer, yang digunakan untuk masalah Klasifikasi dan Regresi. Namun, terutama, ini digunakan untuk masalah Klasifikasi dalam Pembelajaran Mesin. Tujuan dari algoritma SVM adalah untuk membuat garis atau batas keputusan terbaik yang dapat memisahkan ruang n-dimensi ke dalam kelas-kelas sehingga kita dapat dengan mudah menempatkan titik data baru dalam kategori yang benar di masa depan. Batas keputusan terbaik ini disebut *hyperplane*. SVM memilih titik/vektor ekstrim yang membantu dalam menciptakan *hyperplane*. Kasus ekstrim ini disebut sebagai *support vector*, dan karenanya algoritma disebut sebagai *Support Vector Machine*. Pertimbangkan diagram di bawah ini di mana ada dua kategori berbeda yang diklasifikasikan menggunakan batas keputusan atau *hyperplane* [10]:



Gambar 2. kinerja Support vector machine

Naïve bayes

Algoritma Naïve Bayes adalah algoritma pembelajaran terawasi, yang didasarkan pada teorema Bayes dan digunakan untuk memecahkan masalah klasifikasi. Hal ini terutama digunakan dalam klasifikasi teks yang mencakup dataset pelatihan dimensi tinggi. Naïve Bayes Classifier adalah salah satu algoritma Klasifikasi sederhana dan paling efektif yang membantu dalam membangun model pembelajaran mesin cepat yang dapat membuat prediksi cepat. Ini adalah pengklasifikasi probabilistik, yang berarti memprediksi berdasarkan probabilitas suatu objek. Beberapa contoh populer dari Algoritma Nave Bayes adalah penyaringan spam, analisis sentimen, dan pengklasifikasian artikel.

Algoritma Naïve Bayes terdiri dari dua kata Naïve dan Bayes, yang dapat digambarkan sebagai: Nave: Disebut Nave karena mengasumsikan bahwa kemunculan fitur tertentu tidak tergantung pada kemunculan fitur lainnya. Seperti jika buah diidentifikasi berdasarkan warna, bentuk, dan rasa, maka buah merah, bulat, dan manis dikenali sebagai apel. Oleh karena itu setiap fitur secara individual berkontribusi untuk mengidentifikasi bahwa itu adalah sebuah apel tanpa bergantung satu sama lain. Teorema Bayes adalah teorema Bayes juga dikenal sebagai Aturan Bayes atau hukum Bayes, yang digunakan untuk menentukan probabilitas hipotesis dengan pengetahuan sebelumnya. Itu tergantung pada probabilitas bersyarat. Rumus untuk teorema Bayes diberikan sebagai: Algoritma Naïve Bayes Classifier [11]

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Di mana,

$P(A|B)$ adalah Probabilitas posterior: Probabilitas hipotesis A pada kejadian yang diamati B.
 $P(B|A)$ adalah Probabilitas kemungkinan: Probabilitas bukti yang diberikan bahwa probabilitas hipotesis adalah benar

Visualisasi hasil pengujian

Tahap ini untuk menampilkan hasil prediksi, hasil dari model yang digunakan. Jika hasilnya sesuai dengan harapan dapat selesai pada tahap ini, jika belum sesuai hasil bisa dilakukan dengan melihat dataset yang digunakan [12].

Evaluasi

Langkah selanjutnya adalah melakukan evaluasi yang bertujuan untuk mengetahui sejauhmana tingkat akurasi hasil klasifikasi data menggunakan metode naive bayes terhadap data testing/uji.

Hasil dan Pembahasan

Pengumpulan data dilakukan dengan dua tahap, tahap pertama adalah mengumpulkan data yang akan dijadikan data latih/training dan tahap kedua mengumpulkan data yang akan dijadikan data uji/testing. Data yang di ambil adalah komentar masyarakat melalui twitter pada periode januari 2020 - 2022. Jumlah data yang didownload sebanyak 309 dengan 5 atribut. berikut contoh dataset yang terkumpul.

Datetime	Tweet Id	Text	Username	Label
2022-12-26 09:18:46+00:00	1.6073E+18	@ndagels Anjir depan UBP Karawang hahaha jangan jangan gue yg nabrak	FarrasArsenio	1
2021-10-19 17:22:13+00:00	1.45051E+18	Tadi jam 20, Live di TBV Berita kita diskusikan soal Kemiskinan Ekstrem di Karawang bareng pak Sekda, Rektor UBP dan Pengamat Ekonomi. https://t.co/Mm04JJbQjY	kangimantow	1
2021-01-19 06:59:45+00:00	1.35142E+18	@paksasarjana @edukotor Hahaha iyak, anjir rental studio buat ngeband. 💎?💎 ç" suramm. Udah UK jg dong ya? Nadya gak ada rencana lanjut tahun ini? Biar bareng yee gak, gilee UBP Karawang kan... 5 doktor!!!! ;Hë	nurulzirtaf	1
2020-12-04 15:38:47+00:00	1.33488E+18	New post (Enam Tahun Berdiri, Semua Prodi di UBP Karawang Berperingkat B) has been published on Prasasti Jabar - https://t.co/g2iPxybXWp https://t.co/8Euz1z9T97	prasasti_jabar	1
2020-10-08 03:49:03+00:00	1.31405E+18	New post (KKN UBP Karawang Usai, Hasilkan 108 Hak Cipta) has been published on Prasasti Jabar - https://t.co/VYwguhdKvN https://t.co/RvL0o0ZChe	prasasti_jabar	1
2020-09-21 09:12:02+00:00	1.30797E+18	Hallo Twitter, Please do your magic! Aku Sinta, mahasiswa S1 Fak Psikologi UBP Karawang Bantu aku lulus tahun ini ya, dengan cara isi kuesioner aku di link ini. Tolong bantu RT ya kak " 💎:ßHI @jiemiardian https://t.co/65fjcKt8mE https://t.co/4nVwl4gd79	sinchan2993	1
2020-09-18 08:57:53+00:00	1.30688E+18	Hi twitter, please do your magic! Aku sinta, mahasiswa Fakultas Psikologi UBP Karawang ç💎 Mau minta bantuan teman2 untuk isi kuesioner aku. Om bantu retweet ya @ronitoxid Terimakasih https://t.co/65fjcKKJLe	sinchan2993	1
2020-09-16 08:54:32+00:00	1.30615E+18	New post (Tiga Prodi di FTIK UBP Karawang Raih Akreditasi B) has been published on Prasasti Jabar - https://t.co/1QVOdSQzZv https://t.co/6Odh2lqCcv	prasasti_jabar	1

2020-09-15 13:22:54+0 0:00	1.30586E +18	New post (Keren, Belum Lama Berdiri, Prodi Hukum UBP Karawang Terakreditasi 'B') has been published on Prasasti Jabar - https://t.co/8Y9DEpHOLK https://t.co/IRCUGULw6A	prasasti_j abar	1
2020-09-03 09:47:17+0 0:00	1.30146E +18	New post (Tingkatkan Kualitas Pembelajaran, Prodi Ilmu Hukum UBP Karawang Laksanakan Akreditasi) has been published on Prasasti Jabar - https://t.co/mi4g2Mv8qy https://t.co/UZXfRbtmi7	prasasti_j abar	1
2020-09-03 08:43:14+0 0:00	1.30144E +18	New post (PKKMB Daring UBP Karawang Tahap Pertama Diikuti 1.635 Peserta) has been published on Prasasti Jabar - https://t.co/MmfQBWUhmH https://t.co/7wEHE5WbFQ	prasasti_j abar	1

Setelah melakukan pengumpuln data dilakukan langkah langkah *praprocessing* data seperti *cleaning*, *case folding*, *tokenizing*, *stopwords*, *steaming* dan *Tf -idf*. setelah melakukan *praprocessing* data maka didapatkan data seperti dibawah ini.

```

0  {'apes': 0.14285714285714285, 'together': 0.14...
1  {'mahasiswa': 0.045454545454545456, 'universit...
2  {'mahasiswa': 0.047619047619047616, 'universit...
3  {'mahasiswa': 0.045454545454545456, 'universit...
4  {'mahasiswa': 0.047619047619047616, 'universit...
5  {'bahagia': 0.030303030303030304, 'work': 0.03...
6  {'iam': 0.058823529411764705, 'ready': 0.05882...
7  {'benci': 0.14285714285714285, 'universitas': ...
8  {'benci': 0.14285714285714285, 'universitas': ...
9  {'benci': 0.14285714285714285, 'universitas': ...
10 {'benci': 0.14285714285714285, 'universitas': ...

```

Berikut hasil pembobotan kata dengan algoritma tf-idf, untuk menghitung frekuensi setiap kata yang dianggap kata dasar, contoh hasilnya seperti dibawah ini.

term	TF	TF-IDF
benci	0.14285714285714285	0.2593271380911784
universitas	0.14285714285714285	0.10935254887708162
buana	0.14285714285714285	0.12440405111391396
juang	0.14285714285714285	0.12440405111391396
karawang	0.14285714285714285	0.0
in	0.14285714285714285	0.28134866378078677
indonesia	0.14285714285714285	0.28134866378078677

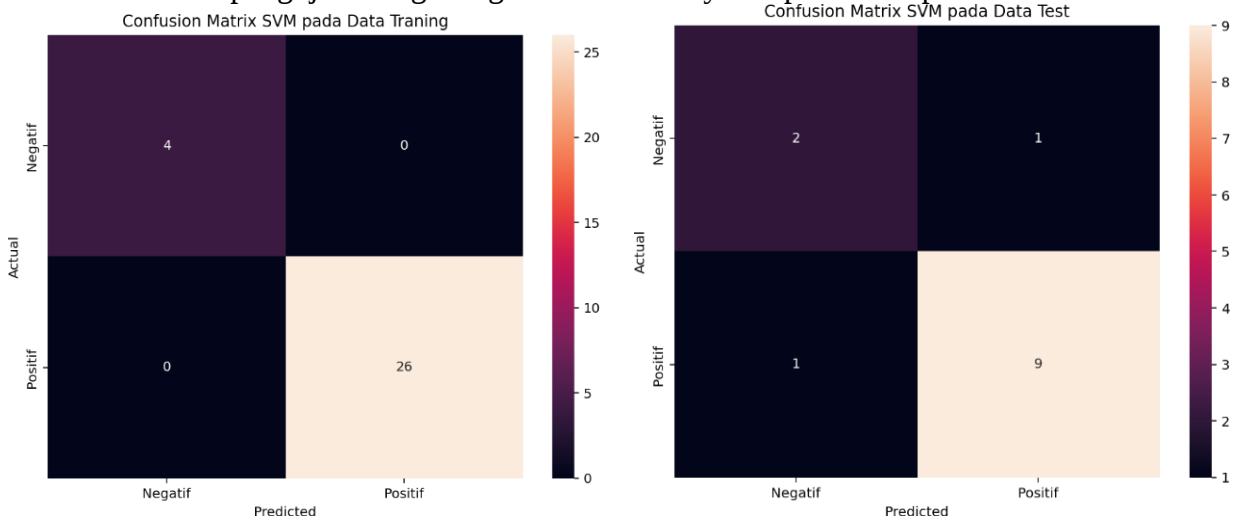
[illegible]

Jumlah dataset yang digunakan pada pengujian menggunakan algoritme support vector machine dan naïve bayes, didapat hasil dari seleksi twett dari masyarakat melalui proses *cleaning, case folding, tokenizing, stopwords, steaming* dan *Tf-idf*. Pada Gambar 4 merupakan grafik data yang digunakan untuk sentiment positif dan negative.

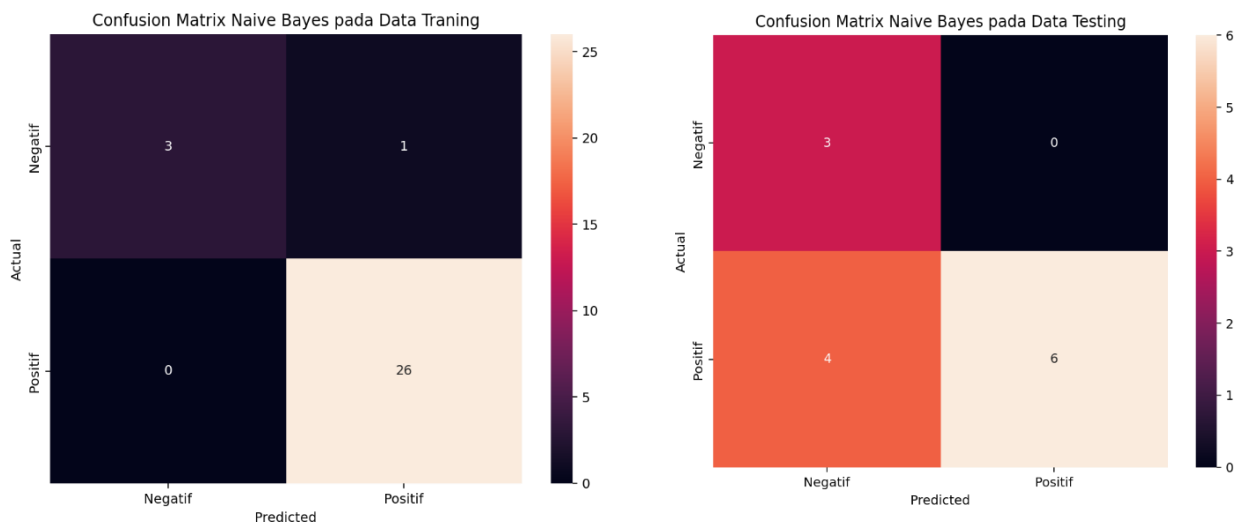


Karawang, 28 Februari 2023

pengujian akurasi pada training dan testing dengan algoritma SVM dapat dilihat pada Gambar 5. Untuk hasil pengujian dengan algoritme naïve bayes dapat dilihat pada Gambar6.



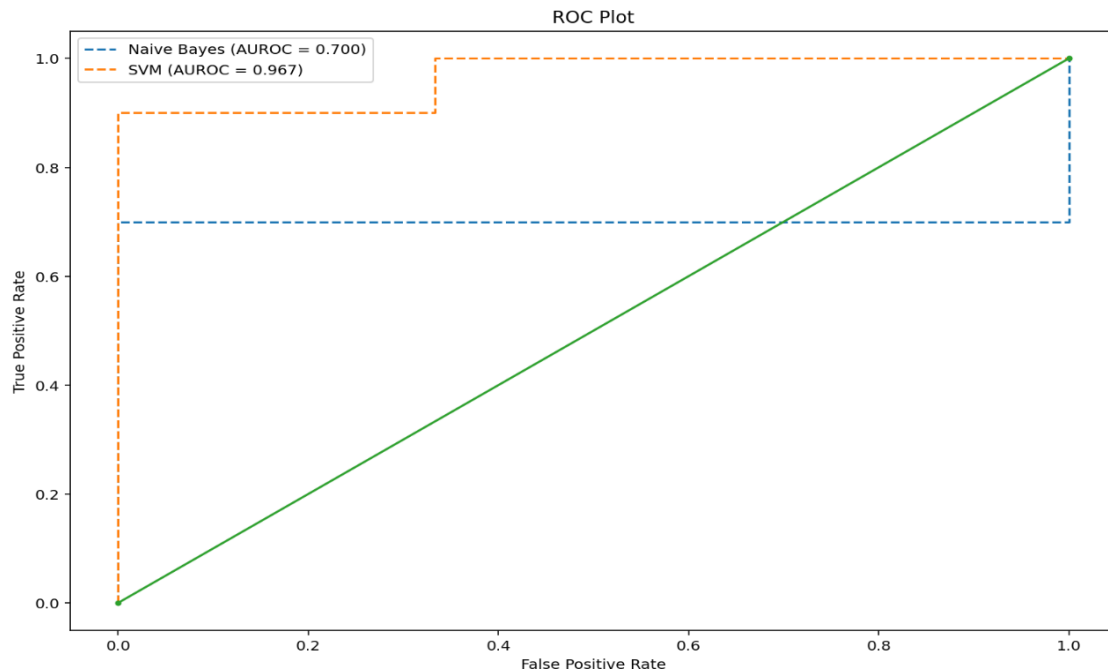
Gambar 5. Hasil confusion matriks SVM



Gambar 6. Hasil confusion matriks Naïve bayes

Jika algoritme yang dilakukan pengujian dengan model akurasi baik support vector machine dan naïve. Hasil akurasi yang diukur adalah akurasi training dan akurasi testing baik algoritma SVM dan Naïve bayes. SVM Accuracy Score Training 86.66 % , SVM Accuracy Score Testing 76.92% dan Naive Bayes Accuracy Score Training 96.66 % , Naive Bayes Accuracy Score Testing 69.23%.

Untuk evaluasi menggunakan ROC dapat di visualkan dalam bentuk gambar, dimana ROC dapat memberikan evaluasi dari perspektif presisian dan recall, sehingga lebih akurat dalam pengambilan keputusan. Hasil ROC dari algoritme yang digunakan dapat dilihat pada Gambar 7. Hasil ROC Naive Bayes 70% dan SVM 96,7%.



Gambar 7. Hasil ROC algoritma SVM dan Naïve Bayes

Pembahasan

Berdasarkan hasil yang dilakukan mulai dari pengambilan data sampai mendapatkan hasil pengujian dapat diambil beberapa poin sebagai perihal yang menarik untuk dapat di diskusikan.

Pengambilan data pada twitter yang didapat tidak banyak, kemungkinan besar kepedulian masyarakat karawang tidak terlalu suka berkomentar dimedia social perihal kehadiran Kampus UBP karawang, dan bisa juga twitter bukan platform media social yang populer di masyarakat karawang.

Dengan dataset yang kurang banyak mengakibatkan model yang didapatkan tidak optimal untuk pembelajaran untuk model tidak cukup data yang didapatkan, sehingga mengalami ketimpangan akurasi training dan testing, dimana hasilnya tidak menunjukkan dataset yang cukup untuk pengujian menggunakan machine learning.

Kesimpulan

1. Berdasarkan hasil data sentimen masyarakat dari twitter terhadap ubp karawang tidak banyak terkumpul sehingga dapat dikatakan sentiment terhadap ubp belum begitu populer.
2. Penerapan *Algoritma Naive Bayes* dan *SVM* telah dilakukan dengan model sentimen analisis dengan machine learning
3. Model akurasi dengan algoritme Naïve Bayes 96.66% dan SVM mencapai 86.66%

Saran

Penelitian berikutnya perlu dilakukan dengan data yang lebih banyak dan mengambil dari platform media social lain seperti Instagram, facebook dan lainnya.

Karawang, 28 Februari 2023

DAFTAR PUSTAKA

- [1] N. D. Putranti and E. Winarko, "Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine," *IJCCS (Indonesian J. Comput. Cybern. Syst.,* vol. 8, no. 1, p. 91, Jan. 2014, doi: 10.22146/ijccs.3499.
- [2] N. A. Vidya, "Twitter Sentimen Analysis Terhadap Brand Reputation," *Univ. Indones.,* 2015.
- [3] M. B. Rozi, I. F., Hamdana, E. N., & Alfahmi, "Pengembangan Aplikasi Analisis Sentimen Twitter Menggunakan Metode Naïve Bayes Classifier (Studi Kasus SAMSAT Kota Malang)," *J. Inform. Polinema,* vol. 4, no. 2, pp. 149–154, 2018.
- [4] E. Mahardhika, Y. S., & Zuliarso, "Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritma Naives Bayes Classifier," *SINTAK,* pp. 409–413, 2018.
- [5] L. Septiani and Y. Sibaroni, "Sentiment Analysis Terhadap Tweet Bernada Sarkasme Berbahasa Indonesia," *J. Linguist. Komputasional,* vol. 2, no. 2, p. 62, Sep. 2019, doi: 10.26418/jlk.v2i2.23.
- [6] S. Hikmawan, A. Pardamean, and S. N. Khasanah, "Sentimen Analisis Publik Terhadap Joko Widodo terhadap wabah Covid-19 menggunakan Metode Machine Learning," *J. Kaji. Ilm.,* vol. 20, no. 2, pp. 167–176, May 2020, doi: 10.31599/jki.v20i2.117.
- [7] B. Liu, "Sentiment Analysis and Opinion Mining. California," *Morgan & Claypool,* 2012.
- [8] K. Dave, S. Lawrence, and D. M. Pennock, "Mining the peanut gallery," in *Proceedings of the twelfth international conference on World Wide Web - WWW '03,* 2003, p. 519, doi: 10.1145/775152.775226.
- [9] A. Nurkholis, D. Alita, and A. Munandar, "Comparison of Kernel Support Vector Machine Multi-Class in PPKM Sentiment Analysis on Twitter," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi),* vol. 6, no. 2, pp. 227–233, Apr. 2022, doi: 10.29207/resti.v6i2.3906.
- [10] S. Styawati and K. Mustofa, "A Support Vector Machine-Firefly Algorithm for Movie Opinion Data Classification," *IJCCS (Indonesian J. Comput. Cybern. Syst.,* vol. 13, no. 3, p. 219, Jul. 2019, doi: 10.22146/ijccs.41302.
- [11] Siregar Amril M; Faisal Sutan; Cahyana Yana; Priyatna Bayu, "Perbandingan Algoritme Klasifikasi Untuk Prediksi Cuaca," *AIMS,* vol. 3, no. 1, 2020, doi: <https://doi.org/10.32627/aims.v3i1.92>.
- [12] R. Tosepu *et al.*, "Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia," *Sci. Total Environ.,* vol. 725, p. 138436, Jul. 2020, doi: 10.1016/j.scitotenv.2020.138436.