

JOINT DISTRIBUTION PADA WEIGHTED MAJORITY VOTE (WMV) UNTUK PENINGKATAN KINERJA SENTIMENT ANALYSIS TERSUPERVISI PADA DATASET TWITTER

Bagus Setya Rintyarna*¹

¹Universitas Muhammadiyah Jember, Jember

Email: ¹bagus.setya@unmuhjember.ac.id

*Penulis Korespondensi

(Naskah masuk: 01 Maret 2022, diterima untuk diterbitkan: 24 Oktober 2022)

Abstrak

Sentiment analysis adalah teknik komputasi *text mining* berbasis *natural language processing* (NLP) untuk mengekstraksi pendapat seseorang yang diungkapkan dalam platform online, termasuk dalam platform *microblogging* Twitter, salah satu platform *microblogging* yang paling populer digunakan di Indonesia. Ada dua pendekatan yang umum digunakan dalam teknik *sentiment analysis* yaitu pendekatan berbasis *machine learning* (ML) dan pendekatan berbasis *sentiment lexicon* (SL). Fokus penelitian ini adalah untuk pengembangan teknik *sentiment analysis* berbasis *machine learning* yang disebut juga teknik tersupervisi pada dataset Twitter. Sebagian besar *sentiment analysis* pada dataset Twitter berbahasa Indonesia mengandalkan *single machine learning algorithm*. Penelitian ini menggabungkan kinerja berbagai algoritma/experts seraya mengurangi tingkat kesalahan klasifikasi dengan meng-update bobot secara dinamis menggunakan *weighted majority vote* (WMV) berbasis *joint distribution* dari Bayesian Network. Pada tahap pertama, data di grabbing dari Twitter dengan 3 hashtag terkait Covid-19 sebagai data eksperimen. Selanjutnya kinerja *weighted majority vote* secara ekstensif dibandingkan dengan 4 metode baseline sebagai pembanding, yaitu: Naïve Bayes, Gaussian Naïve Bayes, Multinomial Naïve Bayes dan Majority Vote dari ketiga *single classifier* tersebut. Metrics kinerja yang digunakan adalah *precision*, *recall*, *fmeasure*, *accuracy* dan *Mathews correlation coefficient* (MCCC). Dalam eksperimen, terbukti bahwa WMV mampu meningkatkan kinerja *sentiment analysis* pada ketiga topik dataset dengan evaluator berbagai metrics kinerja *sentiment analysis*.

Kata kunci: *Sentiment analysis, Machine learning, weighted majority algorithm*

ENHANCING THE PERFORMANCE OF SUPERVISED SENTIMENT ANALYSIS TASK ON INDONESIAN TWITTER DATASET BASED ON WEIGHTED MAJORITY VOTE ALGORITHM

Abstract

Sentiment analysis is a computational *text mining* technique based on *natural language processing* (NLP) to extract someone's opinion expressed in online platforms, including the Twitter *microblogging* platform, one of the most popular *microblogging* platforms used in Indonesia. There are two approaches that are commonly used in *sentiment analysis* techniques, namely the *machine learning* (ML) based approach and the *sentiment lexicon* (SL) based approach. The focus of this research is the development of *machine learning-based* *sentiment analysis* techniques which are also called supervised techniques on the Twitter dataset. Most of the *sentiment analysis* on the Indonesian language Twitter dataset relies on a *single machine learning algorithm*. This study combines the performance of various algorithms/experts while reducing the level of misclassification by updating the weights dynamically using a *joint distribution-based* *weighted majority vote* (WMV) from the Bayesian Network. In the first stage, data was grabbed from Twitter with 3 hashtags related to Covid-19 as experimental data. Furthermore, the performance of the *weighted majority vote* was extensively compared with 4 baseline methods for comparison, namely: Naïve Bayes, Gaussian Naïve Bayes, Multinomial Nave Bayes and Majority Vote from the three *single classifiers*. Performance metrics used are *precision*, *recall*, *fmeasure*, *accuracy* and *Mathews correlation coefficient*. In experiments, it is proven that WMV is able to improve *sentiment analysis* performance on the three dataset topics with various evaluators of *sentiment analysis* performance metrics.

Keywords: *Sentiment analysis, Machine learning, weighted majority vote*

1. PENDAHULUAN (huruf besar, 10pt, tebal)

Perkembangan teknologi internet secara aktif telah mendorong jumlah akses dan interaksi antar penggunaannya melalui berbagai macam platform termasuk social media maupun *microblogging* (Demircan *et al.*, 2021). Kondisi ini sangat menguntungkan dan menjadikan platform online sebagai salah satu sumber berharga untuk mengekstraksi suatu trend sosial (Rintyarna, 2021). Salah satu disiplin yang mendalami topik ini disebut sebagai *computational social science* (CSS), sebuah upaya untuk mengungkapkan fenomena sosial dengan Teknik berbasis komputasi (Rintyarna, Sarno and Faticah, 2018).

Salah satu teknik yang umum digunakan dalam CSS disebut sebagai teknik *sentiment analysis* (SA). *Sentiment analysis* adalah teknik komputasi *text mining* berbasis *natural language processing* (NLP) untuk mengekstraksi pendapat seseorang yang diungkapkan dalam platform online, termasuk dalam platform *microblogging* Twitter, salah satu platform *microblogging* yang paling populer digunakan di Indonesia. Reliabilitas *sentiment analysis* memungkinkan teknik ini mampu mengekstraksi suatu trend sosial tanpa teknik berbasis pencil survey yang membutuhkan waktu yang lama, biaya yang besar dan terkadang menimbulkan bias pada hasilnya (Rintyarna, Sarno and Faticah, 2018).

Ada dua pendekatan yang umum digunakan dalam teknik *sentiment analysis* yaitu pendekatan berbasis *machine learning* (ML) (Rintyarna, Sarno and Faticah, 2019) dan pendekatan berbasis *sentiment lexicon* (SL). *Sentiment analysis* berbasis *machine learning* mengadopsi pendekatan terawasi (*supervised approach*) (Aliyanto, Sarno and Rintyarna, 2017) yang membutuhkan dataset untuk training serta tahap training untuk pemodelan algoritma yang digunakan. Sedangkan pendekatan berbasis *sentiment lexicon* adalah pendekatan yang mengandalkan *sentiment lexicon*, suatu koleksi/kamus/repository digital yang berisi *sentiment word* beserta nilai sentimen kata-kata tersebut (Rintyarna, Sarno and Faticah, 2020).

Fokus penelitian ini adalah untuk pengembangan teknik *sentiment analysis* berbasis *machine learning* pada dataset Twitter. Dataset Twitter telah menarik perhatian banyak penelitian mengingat reliabilitasnya, jumlahnya yang besar serta karakteristiknya yang memungkinkannya menjadi salah satu sumber data penting untuk pemodelan CSS.

Penelitian-penelitian *sentiment analysis* pada dataset Twitter berbahasa Indonesia dengan teknik *machine learning*, umumnya menggunakan *single algorithm* dan beberapa diantaranya dilengkapi dengan optimasi parameter dari algoritma yang digunakan. Bagaimanapun juga, optimasi parameter pada *single classifier* selalu menghadapi trade off berupa “*error classification*” sebagaimana penjelasan (Ruta and Gabrys, 2005). Pada kenyataannya sebagian

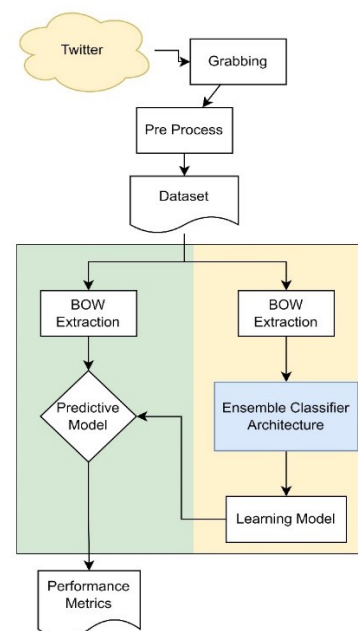
besar *sentiment analysis* pada dataset Twitter berbahasa Indonesia lebih banyak menggunakan *single classifier* dibandingkan *ensemble classifier* (Rintyarna *et al.*, 2022).

Penelitian ini mengusulkan untuk meningkatkan kinerja *sentiment analysis* pada dataset Twitter berbahasa Indonesia berbasis *machine learning* dengan teknik *weighted majority vote* (WMV). WMV mengkombinasikan n predictor untuk menentukan kelas sentiment dataset Twitter. Berbeda dengan majority voting atau weighted ensemble yang memberikan bobot statis pada predictor, WMV mengupdate bobot setiap waktu sehingga mirip dengan perceptron algorithm. Update bobot dilakukan dengan *joint probability distribution* dari Bayesian Network (Bashir, Khan and Awad, 2011). Penelitian ini mengembangkan arsitektur WMV untuk meng-update bobot L expert. Di tahap evaluasi, eksperimen dilakukan menggunakan dataset yang di-grab dari Twitter dengan 3 hashtag terkait Covid-19. Parameter uji berupa metrics kinerja precision, recall, fmeasure, accuracy dan Mathews correlation coefficient.

Selanjutnya, manuskrip ini tersusun atas beberapa bab antara lain Pendahuluan yang membahas latar belakang dan motivasi penelitian, Penelitian terkait yang membahas penelitian *sentiment analysis state of the art* pada dataset Twitter. Bab berikutnya mengulas tentang Metode, Hasil dan Pembahasan serta Kesimpulan.

2. METODE PENELITIAN

Metode penelitian yang digunakan meliputi beberapa tahap aktifitas *machine learning* yaitu grabbing, pre processing, training, testing dan evaluasi sebagaimana digambarkan dalam diagram alir pada Gambar 1.

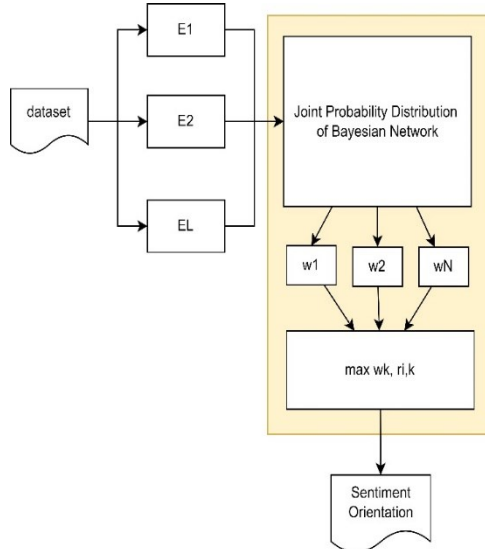


Gambar 1. Diagram alir metode

Pada tahap pertama, dataset untuk eksperimen di-*grabbing* dari Twitter menggunakan Twitter API dengan OAuth2.0 package dari bahasa R. Untuk keperluan filter data pada tahap *grabbing*, digunakan hashtag #covid-19 untuk mengambil Twit yang berkaitan dengan Covid-19, mengambil momentum pandemi yang saat ini sedang terjadi. Jika diperlukan eksperimen dan data tambahan, hashtag yang lain juga memungkinkan untuk digunakan.

Data Twit kemudian di Pre-Process menggunakan tool NLP dari Python sastrawi yang merupakan ekstensi dari PHP Sastrawi (Rintyarna *et al.*, 2022). Pre-process melibatkan proses-proses: (1) *case folding*; (2) *tokenizing*; (3) *normalization*; (4) *stop word removal*; dan (5) *stemming*. Sebagaimana penjelasan (Rintyarna *et al.*, 2022), proses-proses tersebut adalah prosedur standard pada *text pre processing*. Fitur Bag of Word (BoW) kemudian diekstraksi dengan Python sebelum memasuki tahap *Learning* dan *Testing* pada Machine Learning.

Di tahap Learning, optimalisasi classifier dilakukan dengan Arsitektur Ensemble Classifier berbasis Bayesian Network. Arsitektur ensemble classifier yang digunakan dapat dideskripsikan seperti pada Gambar 2. Teknik berbasis ensemble classifier bertujuan untuk meningkatkan performansi classifier dibandingkan kinerja classifier tunggal. Isu penting dalam teknik ini adalah bagaimana ensemble classifier mampu menghasilkan pilihan kelas yang identik dalam hal "*correct classification*". Dan sebaliknya dalam kasus "*error classification*" (Ruta and Gabrys, 2005).



Gambar 2. Arsitektur Ensemble Classifier

Misalnya, ada sejumlah L *expert E* dalam arsitektur *ensemble classifier* untuk input dataset dengan output berupa orientasi sentiment 3 kelas yaitu positive, negative dan netral. Tahap learning dari ensemble classifier bertujuan untuk mencari bobot maksimal w yang menghasilkan orientasi sentiment dengan performance metrics tertinggi. Dapat dinotasikan dalam persamaan (1).

$$c = \max_{k \in C} w_k \sum_{i=1}^L r_{i,k} \tag{1}$$

Di mana c adalah variabel random yang merepresentasikan orientasi sentiment, L adalah jumlah expert dan w adalah bobot yang diestimasi dengan Bayesian Network. Tujuan proses *learning* dari *ensemble classifier*, dalam hal ini adalah untuk menentukan joint probability distribution di mana w_k merepresentasikan bobot yang diestimasi untuk kelas ke- k . Sementara itu, $r_{i,k}$ adalah fungsi yang nilainya 1 jika expert E_i memprediksi sample dataset d pada kelas C_k . Bobot w_k dihitung oleh E_i ketika memenuhi persamaan (2).

$$w_k(e) = p(c = C_k, e) \tag{2}$$

Bayesian Network sendiri sebenarnya menentukan joint probability distribution melalui sebuah struktur *direct acyclic graph* (DAG) di mana node mewakili variabel dan arc mewakili *statistical dependency*-nya. Pada penelitian ini, w_k ditentukan dengan mempertimbangkan Markov Blanket dari node c pada DAG sebagaimana dijelaskan dalam (Heckerman, 2008). Dalam penjelasan (Bashir, Khan and Awad, 2011), *conditional probability distribution* dari variabel random e_i dapat disederhanakan menjadi (3).

$$p(e_i | pa_{e_i}, np_{e_i}) = p(e_i | pa_{e_i}) \tag{3}$$

Di mana pa_{e_i} adalah node dengan parents node e_i . Secara sederhana, joint probability dari seperangkat variabel $\{c, e_1, \dots, e_L\}$ dapat dinyatakan sebagaimana notasi pada persamaan (4).

$$p(c, e_1, \dots, e_L) = p(c | pa_c) \prod_{e_i \in e} p(e_i | pa_{e_i}) \tag{4}$$

Di tahap akhir, untuk mengukur keberhasilan kinerja metode yang diusulkan, hasil diverifikasi dengan performance metrics yaitu precision, recall, f measure dan accuracy berturut-turut yang ditentukan dengan persamaan (5), (6), (7) dan (8).

$$prec = \frac{tp}{tp+fp} \tag{4}$$

$$rec = \frac{tp}{tp+fn} \tag{5}$$

$$fmeas = \frac{2*(rec*prec)}{rec+prec} \tag{6}$$

$$accur = \frac{tp+tn}{tp+tn+fp+fn} \tag{7}$$

Pada persamaan (4) sampai dengan (6), tp menotasikan true positive, tn adalah true negative, fp adalah false positive dan fn adalah *false negative*.

3. PENELITIAN TERKAIT

Penelitian terkait *sentiment analysis* pada dataset Twitter berbahasa Indonesia dengan teknik machine learning umumnya menggunakan *single predictor*. Meskipun, untuk dataset berbahasa Inggris, teknik WMV sudah banyak digunakan, seperti pada penelitian (Aziz and Dimililer, 2020). Beberapa di antaranya dioptimasi parameter dan ditambahkan seleksi fitur untuk peningkatan kinerjanya. Seperti teknik *sentiment analysis* berbasis *machine learning* yang diulas dalam penelitian (Rintyarna, 2017). Penelitian ini menggunakan algoritma Multinomial Naive Bayes yang di-extend menggunakan beberapa teknik text pre-processing. Beberapa teknik *pre-processing* yang digunakan antara lain: 1) IteratedLovinStemmer, 2) LovinStemmer, 3) NullStemmer dan 4) SnowballStemmer. Hasil penelitian diukur berdasarkan metrics kinerja precision, recall dan f-measure. Keempat stemmer tidak berpengaruh signifikan terhadap performansi Multinomial Naive Bayes pada dataset Twitter.

Penelitian lain dilakukan oleh (Mailo and Lazuardi, 2019) dengan Naive Bayes Classifier. Topik yang menjadi obyek penelitian berkaitan dengan permasalahan obesitas di Indonesia. Data di-scraping dari Twitter menggunakan Google Chrome Twitter Testing v.0.1 dalam kurun waktu tahun 2012 sampai dengan 2017. Tool yang digunakan untuk implementasi Naive Bayes Classifier adalah Python 3.7.2 dan R Studio 3.5.2. Dari 43.435 Twit yang digunakan untuk eksperimen setelah proses cleansing, akurasi yang dicapai oleh Naive Bayes dalam Sentiment Analysis adalah sebesar 94%.

Naive Bayes nampaknya adalah classifier yang sering digunakan untuk implementasi sentiment analysis. Seperti yang dilakukan oleh (Ayu Wulandari *et al.*, 2021) pada dataset Twitter. Topik yang diangkat berkaitan dengan persepsi publik terhadap UU Cipta Kerja. Jumlah Twit yang berhasil di crawling mencapai 1667 dataset. Dengan rasio training dan testing sebesar 75:25, eksperimen menggunakan metrics validasi berupa precision, recall dan fmeasure dan accuracy.

Penelitian lain memanfaatkan Sentiment Analysis untuk mempelajari pendapat masyarakat berkaitan dengan pinjaman online (Utami and Erfina, 2021). Data di crawling dari Twitter dengan menggunakan Twitter API. Tidak ada penjelasan tentang berapa jumlah data yang berhasil dicrawling dan digunakan untuk eksperimen. Algoritma yang digunakan berupa algoritma single classifier yaitu Support Vector Machine. Hasil eksperimen mengkonfirmasi bahwa akurasi kinerja Sentiment Analysis pada data pinjaman online melalui Twitter sebesar 62%.

Popularitas teknik komputasi Sentiment Analysis juga pernah digunakan untuk memodelkan pendapat masyarakat berkaitan dengan debat calon presiden Indonesia tahun 2019 (Pratama, Andrian

and Nugroho, 2019). Dengan menggunakan 2 hashtag yaitu #JokowiAminMenangDebat dan #PrabowoIndonesiaMenang, Twitter API berhasil melakukan crawling data sebanyak 15.000 percakapan. Teknik Fine Grained yang diimplementasikan menggunakan Bahasa R dipergunakan untuk memodelkan pendapat masyarakat dengan Teknik Sentiment Analysis.

Teknik sentiment analysis juga pernah dipergunakan untuk memodelkan kecenderungan pendapat masyarakat berkaitan dengan jasa transportasi online yaitu Gojek dan Grab (Pintoko and L., 2018). Dengan menggunakan hashtag #gojekindonesia dan #grab, data di-crawling menggunakan Twitter API. Algoritma single classifier yang dipergunakan sebagai predictor adalah Naive Bayes Classifier. Dengan metrics validator berupa precision, recall dan f-measure, diperoleh informasi bahwa Naive Bayes mampu mencapai rata-rata kinerja sebesar 90%.

Sebuah penelitian tugas akhir melakukan analisis sentiment pada data Twitter yang di-crawling dengan hashtag #bpjsrasatententir (Santoso, 2021). Tujuan penelitian adalah untuk mengetahui opini masyarakat berkaitan dengan kenaikan iuran BPJS pada tahun 2019 sehingga trending hashtag #bpjsrasarententir. Algoritma yang digunakan adalah Support Vector Machine (SVM). Dengan metrics accuracy, kinerja SVM dalam sentiment analysis mencapai 94%.

Dari penelitian-penelitian pendahuluan (sebagaimana tersaji pada Tabel 1) yang dikumpulkan oleh penulis secara intensif melalui study literatur dalam kurun waktu 5 tahun terakhir, sebagian besar teknik tersupervisi (*supervised based technique*) yang dipergunakan untuk Sentiment Analysis pada dataset Twitter adalah teknik berbasis single classifier. Single classifier bagaimanapun memiliki trade off dalam penanganan "*error classification*". Berdasarkan gap tersebut, maka penelitian ini bertujuan untuk mengurangi trade off tersebut dengan menggabungkan beberapa classifier dalam skenario *majority vote*. Namun, teknik *majority vote* yang diadopsi bukan konvensional *majority vote* yang memberikan bobot statis pada tiap prediktor (hanya melihat nilai voting terbesar dari tiap classifier yang digunakan). Tetapi, teknik *majority vote* yang dikembangkan, secara dinamis memberikan bobot pada masing-masing classifier dengan mempertimbangkan *joint probability* pada Bayesian Network antar classifier. Dalam fase validasi, selain menggunakan metrics yang sering digunakan sebagai validator yaitu precision, recall, fmeasure dan accuracy, penelitian ini juga menghitung sebuah metrics yang disebut sebagai Mathews correlation coefficient (MCCC) untuk membuktikan tingkat efektifitas metode yang dikembangkan. Menurut (Chicco and Jurman, 2020), MCCC lebih reliable dibandingkan metrics kinerja lainnya.

Tabel 1. Penelitian "state of the art" terkait

| Penelitian | Dataset | Metode | Jenis |
|--------------------------------------|-----------------------------------|--------------|-------------------|
| (Rintyarna, 2017) | Twitter | Optimasi MNB | Single Classifier |
| (Mailo and Lazuardi, 2019) | Twitter, obesitas | Naive Bayes | Single Classifier |
| (Ayu Wulandari et al., 2021) | Twitter, UU Cipta Kerja | Naive Bayes | Single Classifier |
| (Utami and Erfina, 2021) | Twitter, pinjaman online | SVM | Single Classifier |
| (Pratama, Andrian and Nugroho, 2019) | Twitter, debat calon presiden | Fine Grained | Single Classifier |
| (Pintoko and L., 2018) | Twitter, jasa transportasi online | Naive Bayes | Single Classifier |
| (Santoso, 2021) | Twitter, BPJS | SVM | Single Classifier |

4. HASIL DAN PEMBAHASAN

Jumlah data Twit yang berhasil di grabbing menggunakan API dari OAuth2.0 package adalah sebanyak 19.350 data. Data-data tersebut perlu dianotasi untuk membangun *ground truth* untuk mengevaluasi kinerja *weighted majority vote* berbasis Bayesian Network. Mengingat banyaknya jumlah data yang berhasil di grabbing, anotasi sangat sulit dilakukan secara manual karena membutuhkan banyak waktu dan tenaga. Pada penelitian ini, anotasi dilakukan dengan bantuan package TextBlob pada Tweepy. Hasil anotasi ini selanjutnya digunakan sebagai *ground truth* untuk mengevaluasi kinerja WMV pada metrics kinerja precision, recall, fmeasure dan accuracy. Hasil anotasi ditampilkan pada Tabel 2.

Tabel 2. Hasil anotasi menggunakan TextBlob

| Hashtag | Positive | Negative | Netral |
|-------------|----------|----------|--------|
| #covid19 | 5.647 | 1.254 | 753 |
| #vaksin | 4.289 | 1.134 | 932 |
| #dirumahaja | 3.876 | 987 | 478 |
| Total | 13.812 | 3.375 | 2.163 |

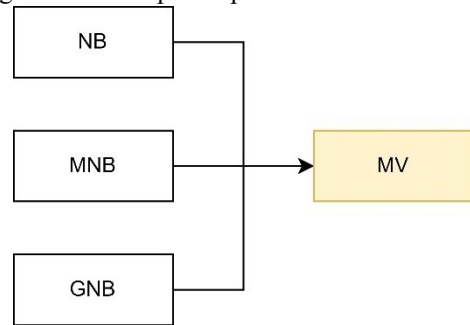
Dari Tabel 2 terlihat bahwa komentar positive masih mendominasi pada ketiga topik dataset disusul dengan komentar negatif dan netral. Untuk mengevaluasi kinerja WMV, digunakan 4 algoritma pembandingan sebagaimana disajikan dalam Tabel 3.

Tabel 3. Algoritma pembandingan dalam eksperimen

| Classifier | Alias | Jenis |
|-------------------------|-------|-------------------|
| Naive Bayes | NB | Single Classifier |
| Gaussian Naive Bayes | GNB | Single Classifier |
| Multinomial Naive Bayes | MNB | Single Classifier |
| Majority Voting | MV | Ensemble |

Untuk teknik Majority Voting (MV) yang digunakan sebagai pembandingan, implementasinya merupakan voting dari NB, MNB dan GNB sebagaimana penjelasan pada Gambar 3. Orientasi

sentiment akhir dari MV dipilih dari voting terbanyak ketiga algoritma pembandingan. Pada Tabel 4, 5 dan 6 berikut ditampilkan hasil eksperimen dengan dataset #covid, #vaksin dan #dirumahaja dengan menggunakan keempat metode pembandingan baik dari jenis Single Classifier maupun Ensemble Classifier sebagaimana ditampilkan pada Tabel 2.



Gambar 3. Teknik Majority Vote sebagai pembandingan

Tabel 4. Hasil eksperimen pada dataset #covid

| | Prec | Rec | F1 | Acc |
|-----|---------------|---------------|---------------|---------------|
| NB | 0.9627 | 0.9268 | 0.9434 | 0.9180 |
| GNB | 0.9614 | 0.9267 | 0.9443 | 0.9194 |
| MNB | 0.9657 | 0.9325 | 0.9488 | 0.9258 |
| MV | 0.9696 | 0.9386 | 0.9538 | 0.9330 |
| WMV | 0.9802 | 0.9738 | 0.9770 | 0.9662 |

Tabel 5. Hasil eksperimen pada dataset #vaksin

| | Prec | Rec | F1 | Acc |
|-----|---------------|---------------|---------------|---------------|
| NB | 0.9357 | 0.8712 | 0.9192 | 0.8928 |
| GNB | 0.9374 | 0.8746 | 0.9205 | 0.8946 |
| MNB | 0.9415 | 0.8824 | 0.9264 | 0.9023 |
| MV | 0.9467 | 0.8925 | 0.9331 | 0.9109 |
| WMV | 0.9613 | 0.9192 | 0.9637 | 0.9509 |

Tabel 6. Hasil eksperimen pada dataset #dirumahaja

| | Prec | Rec | F1 | Acc |
|-----|---------------|---------------|---------------|---------------|
| NB | 0.9614 | 0.926 | 0.9434 | 0.918 |
| GNB | 0.9627 | 0.9267 | 0.9443 | 0.9194 |
| MNB | 0.9657 | 0.9325 | 0.9488 | 0.9258 |
| MV | 0.9696 | 0.9386 | 0.9538 | 0.9330 |
| WMV | 0.9802 | 0.9738 | 0.9770 | 0.9662 |

Pada Tabel 4, 5 dan 6 di atas, performansi terbaik untuk metrics precision, recall, fmeasure dan accuracy diberi tanda dengan notasi bold. Dari eksperimen pada ketiga jenis dataset, tampak bahwa WMV lebih baik dibanding keempat metode pembandingan lain pada keempat metrics kinerja.

Selain keempat metrics performansi, penelitian ini juga menghitung Matthews correlation coefficient (MCC). Dalam penelitian statistik, MCC mirip dengan phi coefficient yang dinotasikan dengan ϕ . Pada *machine learning*, ϕ dikembangkan menjadi MCC untuk mengukur kualitas klasifikasi pada keempat variabel *confusion matrix* yaitu *tp, tn, fp* dan *fn*. MCC dihitung dengan persamaan dan (8).

$$MCC = \frac{tp \times tn - fp \times fn}{\sqrt{(tp+fp)(tp+fn)(tn+fp)(tn+fn)}} \quad (8)$$

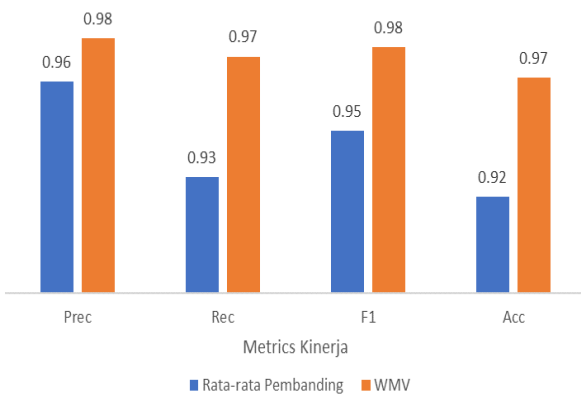
Pada Tabel 6 berikut disajikan hasil perhitungan MCC pada ketiga dataset WMV dibandingkan

dengan semua algoritma pembandingan. Tabel 7 mengindikasikan bahwa kinerja WMV pada metrics MCC juga lebih baik dibandingkan single classifier dan ensemble classifier MV yang melakukan majority voting tanpa bobot.

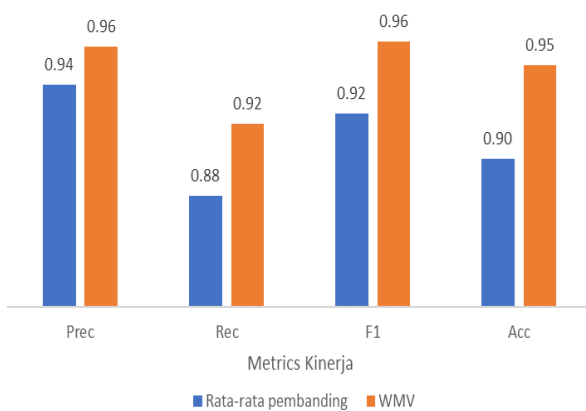
Tabel 7. Hasil perhitungan MCC

| Alg | MCC | | |
|-----|---------------|---------------|---------------|
| | #covid19 | #vaksin | #dirumahaja |
| NB | 0.7966 | 0.7613 | 0.7966 |
| GNB | 0.8004 | 0.7653 | 0.8004 |
| MNB | 0.8156 | 0.7819 | 0.8156 |
| MV | 0.8331 | 0.8008 | 0.8331 |
| WMV | 0.9131 | 0.8878 | 0.9131 |

Untuk mengetahui peningkatan kinerja sentiment analysis dengan menggunakan WMV, dibandingkan dengan dengan rata-rata single classifier dan ensemble classifier yaitu MV, berikut disajikan dalam Gambar 4 dan Gambar 5. Gambar 4 mengindikasikan peningkatan kinerja SA sebesar berturut-turut: 1.5%, 4.3%, 3%, dan 4.2% untuk metrics kinerja precision, recall, fmeasure dan accuracy.



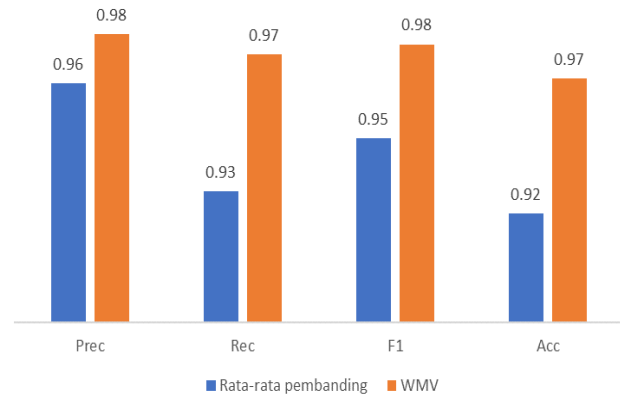
Gambar 4. Peningkatan kinerja SA pada dataset #covid19



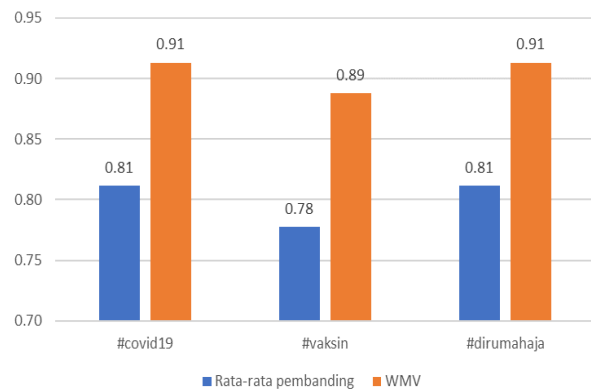
Gambar 5. Peningkatan kinerja SA pada dataset #vaksin

Gambar 5 mengindikasikan peningkatan kinerja SA sebesar: 2.1%, 3.9%, 3.89% dan 5% berturut-turut untuk metrics kinerja precision, recall, fmeasure dan

accuracy. Sedangkan Gambar 6 menghighlight peningkatan kinerja SA sebesar: 1.5%, 4.3%, 2.9% dan 4.2% berturut-turut untuk metrics kinerja precision, recall, fmeasure dan accuracy. Sementara itu, peningkatan MCC tercatat sebesar 10.1%, 11.04%, dan 10.1% berturut-turut pada dataset #covid19, #vaksin dan #dirumahaja sebagaimana tersaji pada Gambar 7.



Gambar 6. Peningkatan kinerja SA pada dataset #dirumahaja



Gambar 7. Peningkatan kinerja MCC pada ketiga dataset

5. KESIMPULAN

Sentiment analysis adalah teknik komputasi text mining berbasis natural language processing (NLP) untuk mengekstraksi pendapat seseorang yang diungkapkan dalam platform online, termasuk dalam platform microblogging Twitter, salah satu platform microblogging yang paling populer digunakan di Indonesia. Penelitian ini mengusulkan untuk meningkatkan kinerja sentiment analysis pada dataset Twitter berbasis machine learning dengan teknik *weighted majority vote* (WMV). WMV mengkombinasikan n predictor untuk menentukan kelas sentiment dataset Twitter. Berbeda dengan majority voting atau weighted ensemble yang memberikan bobot statis pada predictor, WMV mengupdate bobot setiap waktu sehingga mirip dengan perceptron algorithm.

Hasil eksperimen mengkonfirmasi hal-hal sebagai berikut:

1. Pada dataset #covid19, terkonfirmasi peningkatan kinerja SA sebesar berturut-turut: 1.5%, 4.3%, 3%, dan 4.2% untuk metrics kinerja precision, recall, fmeasure dan accuracy.
2. Eksperimen pada dataset #vaksin mengindikasikan peningkatan kinerja SA sebesar: 2.1%, 3.9%, 3.89% dan 5% berturut-turut untuk metrics kinerja precision, recall, fmeasure dan accuracy.
3. Sedangkan eksperimen pada dataset #dirumahaja menghighlight peningkatan kinerja SA sebesar: 1.5%, 4.3%, 2.9% dan 4.2% berturut-turut untuk metrics kinerja precision, recall, fmeasure dan accuracy

6. SARAN

Pada penelitian mendatang, jenis prediktor akan dikembangkan bukan hanya terbatas pada varian Naïve Bayes, dan jenis *probability distribution* dapat diperluas pada model statistik lainnya.

DAFTAR PUSTAKA

- ALIYANTO, D., SARNO, R. & RINTYARNA, B. S., 2017. Supervised Probabilistic Latent Semantic Analysis (sPLSA) for Estimating Technology Readiness Level: International Conference on Information & Communication Technology and System, pp. 79–84.
- AYU WULANDARI, D. dkk, 2021. Analisis Sentimen Media Sosial Twitter Terhadap Reaksi Masyarakat Pada R UU Cipta Kerja Menggunakan Metode Klasifikasi Algoritma Naive Bayes Analysis of Twitter Social Media Sentiment: The Public'S Reaction To the Drafts of Job Creation Law Using the Cla', 8(5), pp. 9007–9016.
- AZIZ, R. H. H. & DIMILILER, N, 2020. Twitter Sentiment Analysis using an Ensemble Weighted Majority Vote Classifier: 3rd International Conference on Advanced Science and Engineering, ICOASE 2020, pp. 103–109. doi: 10.1109/ICOASE51841.2020.9436590.
- BASHIR, A., KHAN, L. & AWAD, M., 2011. Bayesian Networks: Encyclopedia of Data Warehousing and Mining, pp. 89–93. doi: 10.4018/978-1-59140-557-3.ch018.
- CHICCO, D. & JURMAN, G, 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation: BMC Genomics, 21(1), pp. 1–13. doi: 10.1186/s12864-019-6413-7.
- DEMIRCAN, M., dkk., 2021. Developing Turkish Sentiment Analysis Models Using Machine Learning and E-Commerce Data: International Journal of Cognitive Computing in Engineering. doi: 10.1016/j.ijcce.2021.11.003.
- HECKERMAN, D. 2008. A tutorial on learning with Bayesian networks: Studies in Computational Intelligence, 156(November 1996), pp. 33–82. doi: 10.1007/978-3-540-85066-3_3.
- MAILO, F. F. & LAZUARDI, L. 2019. Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia: Journal of Information Systems for Public Health, 4(1), pp. 28–36.
- PINTOKO, B. M. & L., K. M. 2018. Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier: e-Proceeding of Engineering, 5(3), pp. 8121–8130.
- PRATAMA, S. F., ANDREAN, R. & NUGROHO, A. 2019. Analisis Sentimen Twitter Debat Calon Presiden Indonesia Menggunakan Metode Fined-Grained Sentiment Analysis: JOINTECS (Journal of Information Technology and Computer Science), 4(2), p. 39. doi: 10.31328/jointecs.v4i2.1004.
- RINTYARNA, B. S. 2017. Sentiment Analysis pada Data Twitter dengan Pendekatan Naïve Bayes Multinomial: Jurnal Sistem & Teknologi Informasi Indonesia, pp. 1–6.
- RINTYARNA, B. S. 2021. Mapping acceptance of Indonesian organic food consumption under Covid-19 pandemic using Sentiment Analysis of Twitter dataset: Journal of Theoretical and Applied Information Technology, 99(5), pp. 1009–1019.
- RINTYARNA, B. S. dkk. 2022. Modelling Service Quality of Internet Service Providers during COVID-19: The Customer Perspective Based on Twitter Dataset', pp. 1–12.
- RINTYARNA, B. S., SARNO, R. & FATICHAH, C. 2018. Enhancing the performance of sentiment analysis task on product reviews by handling both local and global context: International Journal of Information and Decision Science, 11(xxxx).
- RINTYARNA, B. S., SARNO, R. & FATICHAH, C. 2019. Semantic Features for Optimizing Supervised Approach of Sentiment Analysis on Product Reviews: MDPI Computers, 8(3), pp. 1–16.
- RINTYARNA, B. S., SARNO, R. & FATICHAH, C. 2020. Evaluating the performance of sentence level features and domain sensitive features of product reviews on supervised sentiment analysis tasks: Journal of Big Data, 6(1). doi: 10.1186/s40537-019-0246-8.

- RUTA, D. & GABRYS, B. 2005. Classifier selection for majority voting', *Information Fusion*, 6(1), pp. 63–81. doi: 10.1016/j.inffus.2004.04.008.
- SANTOSO, G. T. 2021. Analisis sentimen pada tweet dengan tagar #bpjsrasarentenir menggunakan metode support vectore machine (svm) skripsi'.
- UTAMI, D. S. & ERFINA, A. 2021. Analisis Sentimen Pinjaman Online di Twitter Menggunakan Algoritma Support Vector Machine (SVM)', *SISMATIK (Seminar Nasional Sistem Informasi dan Manajemen Informatika)*, 1(1), pp. 299–305.