



Analisis Sentimen Berita Online Terhadap Transportasi Online di Indonesia dengan Metode Naïve Bayes Classifier, Support Vector Machine dan K-Nearest Neighbor

Arina Selawati^{1,*}, Yan Rianto², Rachmawati Darma Astuti³, Ainun Zumarniansyah⁴, Deny Novianti⁵

¹ Fakultas Teknik Dan Informatika, Informatika, Universitas Bina Sarana Informatika, Jakarta, Indonesia

² Program Pascasarjana Magister Ilmu Komputer, Universitas Nusa Mandiri, Jakarta, Indonesia

³ Fakultas Teknik Dan Informatika, Teknologi Komputer, Universitas Bina Sarana Informatika, Jakarta, Indonesia

⁴ Fakultas Teknik Dan Informatika, Sistem Informasi, Universitas Bina Sarana Informatika, Pontianak, Indonesia

⁵ Fakultas Teknik Dan Informatika, Sistem Informasi, Universitas Bina Sarana Informatika, Jakarta, Indonesia

Email: ^{1,*}arina.asq@bsi.ac.id, ²yanrianto@gmail.com, ³rachmawati.rcd@bsi.ac.id, ⁴ainun.azm@bsi.ac.id, ⁵deninov.dov@bsi.ac.id

Email Penulis Korespondensi: arina.asq@bsi.ac.id

Abstrak—Berita mengenai transportasi online di Indonesia pada tahun 2019 hingga awal 2020 telah publish di berbagai media online Indonesia, Karena cukup banyak informasi yang ada dalam bentuk teks tanpa ada skala numerik, sulit untuk mengklasifikasikan evaluasi informasi secara efisien tanpa membaca teks secara lengkap. *Sentiment analysis* digunakan untuk mengotomatisasi proses identifikasi pendapat apakah itu adalah pandangan positif atau negatif. Mengklasifikasikan sentimen pada berita dari media berita online dengan proses *Text Mining* dan menggunakan metode *improve Classification Accuracy / Ensemble Methods Techniques* dengan memodifikasi pemodelan algoritma klasifikasi metode *naive bayes, classifier Support vector machine* dan *k-nearest neighbor* yang ditambahkan dengan metode *Particle Swarm Optimization* dan metode *Vote* yang selanjutnya akan dilakukan analisis komparasi. Hasil penelitian diatas adalah mendapatkan nilai akurasi algoritma SVM walau setelah menggunakan fitur seleksi PSO dengan ensembles Vote masih berada di akurasi 84.16%, Begitupun untuk algoritma NB yang mendapatkan 79,08% dan KNN mendapatkan akurasi 87,19%. Kata-kata tersebut akan digunakan untuk melihat kata-kata yang berhubungan dengan sentimen yang sering muncul dan memiliki bobot tertinggi dan dapat digunakan untuk mengetahui artikel berita positif dan artikel berita negatif. Dan untuk penelitian kali ini model yang menggunakan algoritma KNN mendapatkan akurasi tertinggi.

Kata Kunci: Berita Online; Analis Sentimen; Text Mining; Classification Accuracy

Abstract—News about online transportation in Indonesia in 2019 until early 2020 has been published in various Indonesian online media, because there is enough information in the form of text without numerical scale, it is difficult to classify information information efficiently without reading the full text. *Sentiment analysis* is used to automate the process of assessing opinion whether it is positive or negative. Classifying sentiments on news from online news media with the *Text Mining* process and using the method of increasing the *Classification Accuracy / Ensemble Method of Engineering* by combining the classification algorithm *naïve bayes method, classifier Supporting vector machines* and *k-nearest neighbors* added with the *Particle Swarm Optimization* method and *Vote* method The next will be a comparative analysis. The results of the study above get an SVM exam accuracy value even after using the PSO selection feature with the ensemble. Select is still appropriate at 84.16%, Likewise for NB algorithm which gets 79.08% and KNN which gets approval 87.19%. These words will be used to see words related to sentiments that often appear and have the highest weight and can be used to find out positive news articles and negative news articles. And for this research the model that uses KNN algorithm gets the highest accuracy.

Keywords: Online news; Sentiment Analyst; Text Mining; Classification Accuracy

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi saat ini berkembang dengan pesat seiring dengan penemuan dan pengembangan ilmu pengetahuan dalam bidang informasi dan komunikasi sehingga mampu menciptakan alat-alat yang mendukung perkembangan teknologi informasi, mulai dari sistem komunikasi sampai dengan alat komunikasi yang searah maupun dua arah (interaktif)[1]. Sistem informasi memiliki peran yang sangat penting, hal ini terbukti dengan meningkatnya kebutuhan sistem informasi hampir di semua bidang pekerjaan. Dengan adanya perkembangan sistem informasi dalam bidang ilmu pengetahuan, pendidikan, bisnis, administrasi perkantoran, komunikasi, pemerintahan, dan kegiatan lain[2].

Informasi adalah keterangan, pernyataan, gagasan, dan tanda-tanda yang mengandung nilai, makna, dan pesan, baik data, fakta maupun penjelasannya yang dapat dilihat, didengar, dan dibaca, yang disajikan dalam berbagai kemasan dan format sesuai dengan perkembangan teknologi informasi dan komunikasi secara elektronik atau nonelektronik[3]. Website merupakan kumpulan halaman-halaman yang digunakan untuk menampilkan informasi berupa teks, gambar, animasi, suara, ataupun gabungan dari semuanya, baik bersifat statis maupun dinamis yang membentuk suatu rangkaian bangunan yang saling terkait, yang masing-masing dihubungkan dengan jaringan-jaringan halaman[4].

Penyebaran dan perkembangan informasi pada saat ini begitu pesatnya, salah satunya adalah melalui media berita online. Tidak terkecuali di Indonesia dimana sudah banyak media berita online yang memang berasal dari Indonesia dan mengangkat informasi yang terjadi di Indonesia dengan berbagai rubric ataupun kanal di masing-masing medianya. Berita merupakan objek utama sebuah media massa di samping views (opini). Jika dulu, berita hanya disajikan dalam bentuk koran, majalah, dan radio. Pada era millenium generasi net (sekarang) teks berita dengan mudah diperoleh dengan research di google informasi apa yang diinginkan pengguna akan tersedia melalui gawai yang memiliki jaringan internet[5].



Salah satu berita yang ramai diberitakan oleh media berita online pada tahun 2019 hingga awal 2020 adalah berita mengenai transportasi online di Indonesia yang kian berkembang di Indonesia baik motor, mobil, bus, taksi hingga yang akan muncul adalah transportasi online. Berita mengenai transportasi online di Indonesia pada tahun 2019 hingga awal 2020 telah publish di berbagai media online Indonesia antara lain CNN Indonesia, Republika Online, Katadata, BBC Indonesia, Warta Kota, Tribun News, Detik.com, Sindo News, Liputan6.com, Kompas.com, SindoNews.com, tempo.co dan masih banyak lainnya.

Sentiment analysis digunakan untuk mengotomatisasi proses identifikasi pendapat apakah itu adalah pandangan positif atau negatif. Sebuah sistem sentiment analysis otomatis telah dilihat sebagai salah satu alat bisnis intelijen yang diinginkan. Sistem ini dapat mengekstrak opini publik tentang topik tertentu, produk atau jasa yang tertanam dalam teks-teks yang tidak terstruktur. Algoritma klasifikasi sentimen berbasis pembelajaran disisi lain menggunakan proses pembelajaran terhadap data training yang sudah disediakan. Algoritma pembelajaran yang dipakai diantaranya *Naive Bayes* (NB) (He dan Zhou, 2011), *Support Vector Machine* (SVM) (Ghiassi et al., 2013; He dan Zhou, 2011), *Maximum Entropy*(ME) (He dan Zhou, 2011), atau *Fuzzy-NN* (Zhou et al., 2014).

Pada penelitian mengenai sentiment analisis yang sebelumnya dilakukan oleh Nurrin Muchammad Shiddieqy Hadna, Dkk (2016) mengenai studi literatur tentang perbandingan metode untuk proses analisis sentimen di twitter dibahas tahapan yang dilalui untuk melakukan proses analisis sentimen. Dimulai dari tahap *preprocessing* (pelabelan, tokenisasi dan pembobotan), kemudian tahap analisis sentimen dengan perbandingan dua metode yaitu *Naive Bayes Classifier* dan *Support Vector Machine* serta bagaimana mengukur kualitas hasil analisis menggunakan beberapa parameter seperti *accuracy*, *precision* dan *recall*. Atau penelitian yang dilakukan oleh Fachrul Rozy Saputra Rangkuti, Penggunaan seleksi fitur dilakukan untuk meningkatkan performa dari klasifikasi, mengurangi dimensi fitur dan mendapatkan kombinasi fitur yang optimal. Menggunakan beberapa model dari klasifikasi *Naive Bayes* yaitu Bernoulli *Naive Bayes* untuk tipe data biner, *Gaussian Naive Bayes* untuk tipe data kontinu dan *Multinomial Naive Bayes* untuk tipe data numeric. Dengan Hasil penelitian menunjukkan bahwa dengan menggunakan kata tidak baku pada *tweet* hasil evaluasi yang didapatkan *accuracy* 82%, *precision* 86%, *recall* 79,62% dan *f-measure* 82,69% pada seleksi Fitur 20%. Kemudian setelah dilakukan pembakuan kata secara manual hasil evaluasi pada *accuracy* meningkat sebesar 8% sehingga *accuracy* menjadi 90%, *precision* 92%, *recall* 88,46% dan *f-measure* 90,19% pada seleksi fitur 85%. Berdasarkan hasil tersebut dapat disimpulkan bahwa penggunaan kata baku dapat meningkatkan performa klasifikasi dan seleksi fitur Pearson's memberikan kombinasi fitur yang optimal dan mengurangi jumlah total dimensi fitur.

Dari latar belakang dan berbagai referensi penulis dala penelitian ini akan melakukan proses analisa sentimen terhadap berita dengan kata kunci transportasi online di Indonesia. Penelitian ini akan melakukan *improve Classification Accuracy / Ensemble Methods Techniques* dengan memodifikasi pemodelan algoritma klasifikasi metode *naive bayes*, *classifier Support vector machine* dan *k-nearest neighbor* yang ditambahkan dengan metode *Particle Swarm Optimization* dan metode *Vote* yang selanjutnya akan dilakukan analisis komparasi terhadap pemodelan tersebut untuk prediksi terhadap berita transportasi online di Indonesia.

2. METODOLOGI PENELITIAN

Data yang akan penulis gunakan merupakan data yang berhasil penulis kumpulkan secara manual dari media berita online di Indonesia. Penulis mencari judul berita online di media online dengan kata kunci pencarian (*keyword*) yaitu transportasi online di Indonesia. Dari banyaknya jumlah berita online tentang transportasi online di Indonesia, penulis mengumpulkan data berupa judul berita online tentang transportasi online di Indonesia, yang dimuai dari awa tahun 2019 hingga awal tahun2020 (1 tahun).

Data yang terkumpul sejumlah 1023 judul berita, yang penulis akan gunakan sebagai objek penelitian untuk mengklasifikasikan berita berita positif dan berita *negative*. Teks yang belum diolah biasanya memiliki karakteristik dimensi yang tinggi, terdapat noise pada data dan terdapat struktur teks yang tidak baik. Untuk itu, dalam pengolahan data awal, text mining harus melalui beberapa tahapan yang disebut dengan preprocessing.

a. Tokenize

Tokenize disebut sebagai Proses untuk pemotongan kata dalam suatu kalimat kedalam bentuk token, dimana tiap kata dalam satu kalimat dipisahkan oleh spasi[6]. Tokenisasi juga dapat di artikan sebagai proses memecah kalimat menjadi kata-kata,ptanda baca, dan ekspresi bermakna lainnyapsesuai dengan aturan bahasa[7].

b. Filter Token (By Length)

Filter tokens by length digunakan untuk menghapus kata dengan jumlah karakter yang kurang dari nilai yang ditentukan[8]. Proses ini akan memfilter token berdasarkan batas minimum dan maksimum yang telah ditentukan yaitu batas minimum yang dipakai adalah 4 karakter dan 25 karakter untuk batas maksimum karakter[9].

c. Stopwords Removal

menghapus kata tertentu yang tidak dibutuhkan. Stopword removal ini menggunakan kamus bahasa indonesia bersumber dari situs kaggle yang ditambahkan kembali kata stopwords yang relevan berdasarkan data yang diteliti[10]. stopwords removal digunakan untuk menghilangkan kata tidak penting dan tidak memiliki arti seperti kata "dan", "yang", "atau", filtering juga dapat mengurangi dimensi terhdap data text input, sehingga proses akan lebih mudah dijalankan[11].

d. Transform Cases

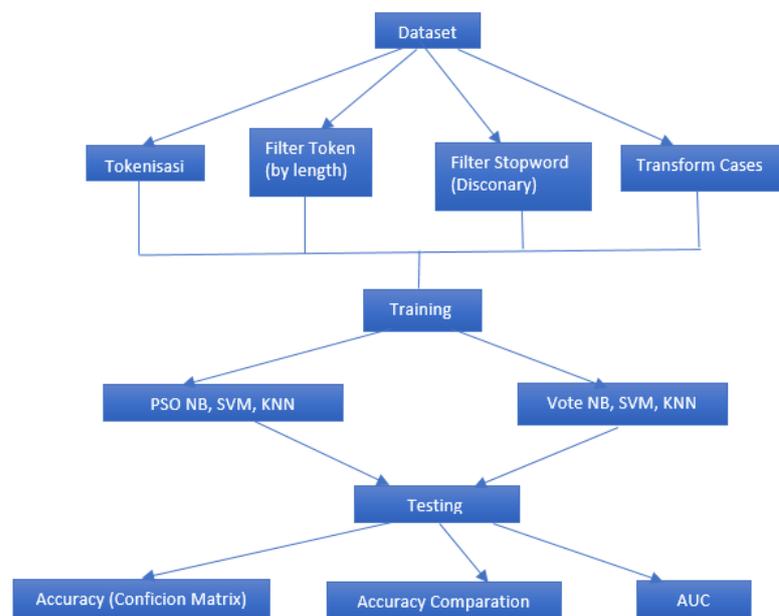


Transform Cases yaitu mengganti karakter huruf yang awalnya huruf besar menjadi huruf kecil, karakter huruf pada text juga berpengaruh terhadap sifat text tersebut[12]. Transform Cases adalah tahap mengubah semua huruf kapital menjadi huruf kecil[13].

Metode atau model yang diusulkan oleh penulis yaitu menggunakan metode *improve Classification Accuracy / Ensemble Methods Techniques* dengan memodifikasi pemodelan algoritma klasifikasi metode *naïve bayes*, *classifier Support vector machine* dan *k-nearest neighbor* yang ditambahkan dengan metode *Particle Swarm Optimization* dan metode *Vote* yang selanjutnya akan dilakukan analisis komparasi. Selain itu metode yang diusulkan menggunakan metode CRISP-DM yang bertujuan untuk memperoleh suatu pola atau pengetahuan dari data yang dimanfaatkan untuk menyelesaikan suatu masalah dan digunakan untuk pengambilan keputusan yang tepat. CRISP-DM adalah metode yang menyediakan standar baku pada data mining dan dapat diterapkan pada strategi pemecahan masalah umum[14]. Cross Industry Standar Process for Data Mining (CRISP-DM) yang terdiri dari enam tahapan yakni tahap business understanding, tahap data understanding, tahap data preparation, tahap modeling, tahap evaluation dan tahap deployment[15].

Dalam melakukan pengujian model *software* yang digunakan untuk mengolah data dan alat bantu untuk menghitung tingkat akurasi, penulis menggunakan aplikasi *RapidMiner versi 9.5*. *Rapidminer* adalah sebuah tools atau software untuk mempermudah pengolahan data, rapid miner tersedia dalam bentuk software free to use atau gratis jika pemakaiannya untuk kepentingan edukasi atau pembelajaran[16]. Rapid Miner adalah software pengolah data. Menggunakan prinsip dan algoritme data mining, RapidMiner mengekstrak pola dari big data, menggabungkan metode statistik, kecerdasan buatan, dan database[17]. Dataset yang digunakan untuk pengujian model diperoleh dari judul artikel berita tentang transportasi online di Indonesia sebanyak 1000 data, yang kemudian diklasifikasikan kedalam berita positif dan berita negatif.

Pada tahap ini diusulkan model untuk data berita online tentang transportasi online di indonesia dengan menerapkan algoritma klasifikasi. Algoritma klasifikasi dapat menggunakan matriks konfusi sebagai salah satu sarana untuk mengevaluasi akurasi algoritma[18] dengan metode *naïve bayes* merupakan salah satu algoritma klasifikasi yang banyak digunakan pada Data Mining ataupun Text Mining. Algoritma Naive Bayes berdasarkan teorema Bayes bahwa semua kegiatan memberikan sebuah kontribusi yang sama penting atau saling bebas pada pemilihan kelas tertentu[19]. *classifier Support vector machine* bertujuan untuk menemukan fungsi pemisah terbaik di antara fungsi yang ada untuk memisahkan dua macam obyek[20] dan *k-nearest neighbor* adalah sejenis *classifier centroid* terdekat. Ide utama algoritma K-NN pertama, hitung jarak atau kesamaan antara sampel uji dan semua sampel pelatihan, dan temukan tetangga terdekat K dari sampel uji dalam sampel pelatihan; kemudian sesuai dengan kategori tetangga ini untuk menentukan kategori sampel uji[21]. Yang ditambahkan dengan metode *Particle Swarm Optimization* dan metode *Vote* untuk meningkatkan akurasi pada hasil perhitungan. Dalam penelitian ini penulis juga menggunakan validasi standar yaitu *10 fold cross-validation* dimana proses ini membagi data secara acak ke dalam 10 bagian. Proses yang dilakukan setelah melakukan pengujian yaitu mengukur *performance* dari algoritma klasifikasi *text mining* yang dipakai. Dalam penelitian ini *performance* diukur menggunakan *Accuracy* dan *AUC* serta akan ditampilkan dalam bentuk kurva *ROC*. Berikut ini adalah alur penelitian yang akan dilakukan yaitu:



Gambar 1. Alur Penelitian

Dari tahapan diatas maka akan didapatkan hasil komparasi antara algoritma: *Naive Bayes*, *Support Vector Machine*, *k-nearest neighbor*, *Naive Bayes (PSO)*, *Support Vector Machine (PSO)* dan *k-nearest neighbor (PSO)* dan hasil Ensembles dari tiga algoritma *Naive Bayes*, *Support Vector Machine*, *k-nearest neighbor*.



3. HASIL DAN PEMBAHASAN

Pada tahapan *business understanding*, dilakukan pemahaman terhadap objek penelitian. Pemahaman mengenai objek penelitian dilakukan dengan menggali informasi melalui beberapa situs berita *online* terhadap transportasi online di Indonesia. Motivasi pada fase ini yaitu berita yang disajikan biasanya dalam bentuk teks pada media digital yang dikelompokkan berdasarkan judul pembahasan dari masing-masing kategori berita. Analisis sentimen ini dilakukan untuk mencari metode klasifikasi yang dapat membantu dalam menentukan judul artikel berita masuk kedalam dua status positif dan negatif.

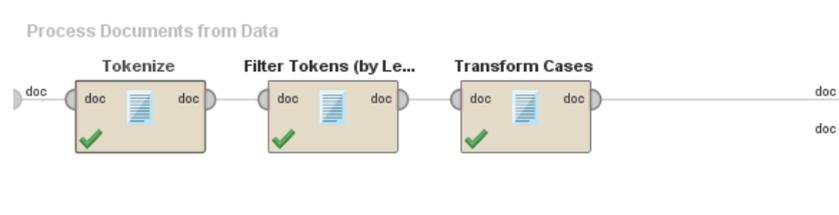
Berikut ini adalah sebagian data berita bertatus positif dan berstatus negatif dari berita tentang transportasi online di Indonesia yang telah penulis kumpulkan melalui *website* media berita *online* sebagai berikut:

Tabel 1. Tabel Contoh Status Berita *Online* Positif dan Negatif

| No | Text | Status |
|----|---|---------|
| 1 | Kemenhub Dukung Operator Online Gunakan Kendaraan Listrik | Positif |
| 2 | Para Penantang Gojek dan Grab Di Bisnis Ojek Online | Negatif |
| 3 | 3 Aplikasi Ojek Online Terbaru di Indonesia Saingi Goje dan Grab, Ada yang Viral | Negatif |
| 4 | 20 Unit Mobil Listrik Grab-Hyundai Mengaspal di Awal 2020 | Positif |
| 5 | Tangis Tukang Gojek Online dapat Tergadaikan | Negatif |
| 6 | Sosialisasi Penerapan PM 118/2018 Dukung Transportasi Aman | Positif |
| 7 | Zona Merah Ojek Online Di Hapus | Positif |
| 8 | Streaming! Laju Industri Otomotif Indonesia di 2020 | Positif |
| 9 | Helmnya Bersayap dan Berjaket Merah, Gaspol Jek Ramaikan Transportasi Online di Indonesia | Positif |
| 10 | Medan Pertarungan Grab dan Gojek Pindah Ke Malaysia | Negatif |
| 11 | Mobil Listrik Ini Mulai Diuji Sebagai Armada Taksi Online | Positif |
| 12 | Ojol: Massa Pengemudi Ojek Online Bawa Paksa Jenazah Bayi, Sejauh Mana Kekuatan Mereka bisa dikerahkan? | Negatif |
| 13 | Grab Ingin Gunakan Armada Kendaraan Listrik di Seluruh Dunia | Positif |
| 14 | Cerita Di Balik Lahirnya Aturan Taksi Online dan Ojol RI | Positif |
| 15 | Terungkap Empat Alasan Ojek Online Tidak Laku di Jepang | Negatif |
| 16 | Gojek GoPay Jalin Kerja Sama Dengan Pemkot Sukabumi | Positif |
| 17 | Grab Latah, Siapkan Armada Mobil Listrik Di Indonesia | Negatif |
| 18 | Ojek Online Dimata Transportasi Publik | Positif |
| 19 | Layanan Beyond Taxi dan Jurus Blue Bird hadapi Badai Disrupsi Transportasi Online | Positif |
| 20 | Dst.... | |

Pada tahap ini juga dilakukan pemahaman untuk mencari metode klasifikasi yang terbaik agar dapat membantu pada saat proses pengolahan data yang akan dilakukan dengan cara membandingkan hasil dari algoritma yang digunakan dan untuk meningkatkan performa dari metode klasifikasi dapat dilakukan dengan menggunakan *feature selection* dan teknik *ensembles (vote)*.

Tahap *data preparation* merupakan tahap dengan proses penyiapan data yang bertujuan untuk mendapatkan data yang bersih dan siap untuk digunakan dalam penelitian. Dalam *text mining* tahapan awal yang akan dilakukan adalah tahap *preprocessing*. Berikut merupakan tahapan yang dilakukan dalam *preprocessing*. Berikut merupakan tahap desain model *preprocessing* yang digunakan untuk tahapan awal pengolahan data :



Gambar 2. Desain Model *Preprocessing*

Dalam proses tokenisasi, semua kata yang ada di dalam tiap dokumen akan dikumpulkan dan dihilangkan tanda baca, simbol, karakter khusus ataupun bukan huruf. dikarenakan dataset yang penulis miliki sudah cukup bersih maka teks sebelum dan sesudah tidak ada perubahan.

Tabel 2.Perbandingan Teks Sebelum dan Sesudah Dilakukan Proses *Tokenize*

| Proses | Teks |
|--|---|
| Teks sebelum dilakukan tokenisasi | kemenhub dukung operator online kendara listrik |
| Teks sesudah dilakukan proses tokenisasi | kemenhub dukung operator online kendara listrik |



Dalam proses filter token (*By Length*) ini, kata-kata yang memiliki panjang kurang dari 4 atau lebih dari 25 akan dihapus, seperti kata yg, tdk, jd, ga, ane, gan yang merupakan kata-kata yang tidak mempunyai makna tersendiri jika dipisahkan dengan kata lain dan tidak terkait dengan kata sifat yang berhubungan dengan sentimen.

Tabel 3.Perbandingan Teks Sebelum dan Sesudah Dilakukan Proses *filter token (By Length)*

| Proses | Teks |
|---|---|
| Teks sebelum dilakukan proses <i>filter token (by length)</i> | kemenhub dukung operator online kendaraan listrik |
| Teks sesudah dilakukan proses <i>filter token (by length)</i> | kemenhub dukung operator online kendaraan listrik |

Pada tahap ini, penulis menggunakan *filter stopwords (dictionary)* karena dataset yang digunakan berbahasa Indonesia. Pada proses ini terlebih dahulu dibuat daftar kata-kata yang termasuk *stopwords* kemudian *file* nya akan *diupload* ke dalam operator *filter stopwords (dictionary)*. Dalam tahap ini, kata-kata yang tidak relevan akan dihapus seperti kata tetapi, untuk, dengan, yang merupakan kata-kata yang tidak memiliki makna tersendiri jika dipisahkan dengan kata yang lain dan tidak terkait dengan kata sifat yang berhubungan sentimen.

Tabel 4.Perbandingan Teks Sebelum dan Sesudah Dilakukan Proses *stopword removal*

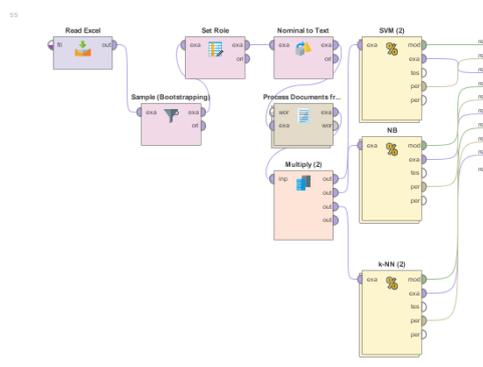
| Proses | Teks |
|--|---|
| Teks sebelum dilakukan proses <i>Stopwords Removal</i> | kemenhub dukung operator online kendaraan listrik |
| Teks sesudah dilakukan proses <i>Stopwords Removal</i> | kemenhub dukung operator online kendaraan listrik |

Dalam proses *transform cases* kata yang mengandung huruf besar diubah menjadi huruf kecil sehingga dapat berhubungan dengan sentimen.

Tabel 5.Perbandingan Teks Sebelum dan Sesudah Dilakukan Proses *Transform Cases*

| Proses | Teks |
|--|---|
| Teks sebelum dilakukan proses <i>Transform Cases</i> | kemenhub dukung operator online kendaraan listrik |
| Teks sesudah dilakukan proses <i>Transform Cases</i> | Kemenhub dukung operator online kendaraan listrik |

Fase Permodelan Adalah fase pemilihan teknik mining dengan menentukan algoritma yang akan digunakan. *Tool* yang digunakan adalah *RapidMiner* versi 9.5. Hasil pengujian model yang dilakukan adalah mengklasifikasikan artikel berita positif dan artikel berita negatif menggunakan algoritma *naïve bayes*, *classifier Support vector machine* dan *k-nearest neighbor*, berikut adalah desain model *Rapidminer* yang digunakan yaitu:



Gambar 3. Desain Model Perbandingan Tiga Algoritma

Adapun perbandingan hasil komparasi akurasi dan AUC Algoritma telah digunakan sebagai berikut:

Tabel 6.Perbandingan Akurasi dan AUC

| Algoritma | Accuracy | AUC |
|------------|----------|-------|
| SVM | 84,16% | 0.919 |
| SVM + PSO | 84,16% | 0.919 |
| SVM + VOTE | 84,16% | 0.919 |
| NB | 79,08% | 0.684 |
| NB + PSO | 79,08% | 0.684 |
| NB + VOTE | 79,08% | 0.684 |
| KNN | 87,19% | 0.500 |
| KNN+PSO | 87,19% | 0.500 |
| KNN+VOTE | 87,19% | 0.500 |



Hasil penelitian diatas adalah mendapatkan nilai akurasi algoritma SVM walau setelah menggunakan fitur seleksi PSO dengan ensembles Vote masih berada di akurasi 84.16%, Begitupun untuk algoritma NB yang mendapatkan 79,08% dan KNN mendapatkan akurasi 87,19%. Dan Model klasifikasi teks yang digunakan dapat memudahkan untuk mengetahui judul artikel berita positif dan artikel berita negatif. Berdasarkan data berita yang diolah menggunakan *tool Rapidminer*, data artikel berita akan terpisah menjadi kata-kata yang memiliki bobot pada setiap kata-katanya. Kata-kata tersebut akan digunakan untuk melihat kata-kata yang berhubungan dengan sentimen yang sering muncul dan memiliki bobot tertinggi dan dapat digunakan untuk mengetahui artikel berita positif dan artikel berita negatif. Dan untuk penelitian kali ini model yang menggunakan algoritma KNN mendapatkan akurasi tertinggi.

4. KESIMPULAN

Dari hasil penelitian yang penulis lakukan mengenai analisis sentimen judul berita transportasi online, penulis dapat mengambil kesimpulan hasil perhitungan ketiga metode klasifikasi yakni SVM, NB dan KNN yang mendapatkan akurasi tertinggi adalah algoritma KNN dengan akurasi 87,19% . Untuk penambahan pada ketiga algoritma dalam penelitian ini tidak membawa perubahan ataupun menaikkan akurasi dari algoritmanya. Model klasifikasi teks yang digunakan dapat memudahkan untuk mengetahui judul artikel berita positif dan artikel berita negatif. Berdasarkan data berita yang diolah menggunakan *tool Rapidminer*, data artikel berita akan terpisah menjadi kata-kata yang memiliki bobot pada setiap kata-katanya. Kata-kata tersebut akan digunakan untuk melihat kata-kata yang berhubungan dengan sentimen yang sering muncul dan memiliki bobot tertinggi dan dapat digunakan untuk mengetahui artikel berita positif dan artikel berita negatif. Untuk penelitian kali ini model yang menggunakan algoritma KNN mendapatkan akurasi tertinggi sehingga dapat disimpulkan bahwa algoritma KNN menjadi algoritma yang terbaik di penelitian kali ini, walaupun KNN dikenal sebagai algoritma yang sederhana tapi tetap efektif untuk masalah klasifikasi selain itu KNN menjadi salah satu algoritma yang mudah diimplementasikan dan tidak membutuhkan pelatihan model. Saran-saran yang ingin penulis sampaikan untuk penelitian selanjutnya adalah menggunakan pengklasifikasian lain sehingga bisa dilakukan penelitian yang berbeda dari yang sudah ada, menggunakan metode pemilihan fitur yang lain seperti *Chi Square*, *Mutual Information*, *unigram*, *unigram + bigram*, *unigram + Part of Speech (POS)*, dan lain-lain agar hasilnya bisa dibandingkan, dan dapat dilakukan penelitian lebih lanjut terkait dengan tahapan *preprocessing* yang lebih detail yang dikaitkan dengan metode klasifikasi.

REFERENCES

- [1] M. S. I. Lubis, "Teknologi Informasi Dan Komunikasi Dalam Perspektif Islam," in *Prosiding*, 2021, pp. 79–88.
- [2] N. H. Elvi Rahmi, Eva Yumami, "Analisis Metode Pengembangan Sistem Informasi Berbasis Website : Systematic Literature Review," *Remik Ris. dan E-Jurnal Manaj. Inform. Komput.*, vol. 7, pp. 821–834, 2023.
- [3] Mukhsin, "Peranan Teknologi Informasi Dan Komunikasi Dalam Publikasi Informasi Desa Di Era Globalisasi," *TEKNOKOM*, vol. 3, no. 1, pp. 7–15, 2020.
- [4] M. A. K. Rizki and A. F. OP, "Rancang Bangun Aplikasi E-Cuti Pegawai Berbasis Website (Studi Kasus : Pengadilan Tata Usaha Negara)," *J. Teknol. dan Sist. Informaasi*, vol. 2, no. 3, pp. 1–13, 2021.
- [5] I. Riana, Y. Sinambela, M. Surip, and S. F. Dalimunthe, "Analisis Wacana Kritis pada Berita Online Crazy Rich Indonesia," *J. Pendidik. Tambusai*, vol. 6, no. November 2021, pp. 9421–9429, 2022.
- [6] N. S. Marga, A. R. Isnain, and D. Alita, "Sentimen Analisis Tentang Kebijakan Pemerintah Terhadap Kasus Corona Menggunakan Metode Naive Bayes," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 2, no. 4, pp. 453–463, 2020.
- [7] B. Kurniawan *et al.*, "Sentimen Analisis Terhadap Kebijakan Penyelenggara Sistem Elektronik (Pse) Menggunakan Algoritma Bidirectional Encoder Representations From Transformers (Bert)," *J. Teknol. dan Sist. Informaasi*, vol. 3, no. 4, pp. 98–106, 2022.
- [8] A. P. Giovani, T. Haryanti, and L. Kurniawati, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *J. Teknoinfo*, vol. 14, no. 2, pp. 116–124, 2020.
- [9] A. Hendra and Fitriyani, "Analisis Sentimen Review Halodoc Menggunakan Nai" ve Bayes Classifier," *JISKA*, vol. 6, no. 2, pp. 78–89, 2021.
- [10] Irvandi, B. Irawan, and O. Nurdiawan, "Naive Bayes Dan Wordcloud Untuk Analisis Sentimen Wisata Halal Pulau Lombok," *INFOTECH*, vol. 9, no. 1, pp. 236–242, 2023.
- [11] A. R. Isnain, A. I. Sakti, D. Alita, and N. S. Marga, "Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan algoritma Svm," *JDMISI*, vol. 2, no. 1, pp. 31–37, 2021.
- [12] Surohman, S. Aji, Rousyanti, and F. F. Wati, "Analisa Sentimen Terhadap Review Fintech Dengan Metode Naive Bayes," *J. Sains dan Manaj.*, vol. 8, no. 1, pp. 93–105, 2020.
- [13] M. I. Ahmadi, D. Gustian, and F. Sembiring, "Analisis Sentiment Masyarakat terhadap Kasus Covid-19 pada Media Sosial Youtube dengan Metode Naive bayes," *J. Sains Komput. Inform.*, vol. 5, no. September, pp. 807–814, 2021.
- [14] J. W. Iskandar and Y. Nataliani, "Perbandingan Naive Bayes, SVM, dan K-NN Untuk Analisis Sentimen Gadget Berbasis Aspek," *J. RESTI*, vol. 5, no. 158, pp. 1120–1126, 2021.
- [15] Y. A. Singgalen, "Analisis Perilaku Wisatawan Berdasarkan Data Ulasan di Website Tripadvisor Menggunakan CRISP-DM : Wisata Minat Khusus Pendakian Gunung Rinjani dan Gunung Bromo," *J. Comput. Syst. Informatics*, vol. 4, no. 2, pp. 326–338, 2023, doi: 10.47065/josyc.v4i2.3042.
- [16] M. R. Nahjan, N. Heryana, and A. Voutama, "Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell," *J. Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 101–104, 2023.
- [17] Ainurrohman, "Akurasi Algoritma Klasifikasi pada Software Rapidminer dan Weka," *PRISMA*, vol. 4, pp. 493–499, 2021.



- [18] D. Pascalina, R. Widhiastono, and C. Juliane, “Pengukuran Kesiapan Transformasi Digital Smart City Menggunakan Aplikasi Rapid Miner,” *Technomedia Journal*, vol. 7, no. 3, pp. 293–302, 2023.
- [19] D. Darwis, N. Siskawati, and Z. Abidin, “Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional,” *Jurnal Tekno Kompak*, vol. 15, no. 1, pp. 131–145, 2021.
- [20] N. Hendrastuty, A. R. Isnain, and A. Y. Rahmadhani, “Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine,” *J. Pengemb. IT*, vol. 6, no. 3, pp. 150–155, 2021.
- [21] M. S. Alrajak, I. Ernawati, and I. Nurlaili, “Analisis Sentimen Terhadap Pelayanan Pt Pln Di Jakarta Pada Twitter Dengan Algoritma K- Nearest,” *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, pp. 110–122, 2020.