

Visualisasi Data Analisa Sentimen RUU *Omnibus law* Kesehatan Menggunakan KNN dengan Software RapidMiner

Tupari^{1*)}, Syaukani Abdullah², Chairani³

^{1,2,3} Fakultas Ilmu Komputer, Magister Teknik Informatika, Institut Informatika dan Bisnis Darmajaya, Bandar Lampung

^{1,2,3} Jalan Z.A. Pagar Alam, No.93 Kedaton, Bandar Lampung, Lampung, 34152, Indonesia

E-mail : ¹tupari.2121210017@mail.darmajaya.ac.id, ²syaukani.2121210004@mail.darmajaya.ac.id,

³chairani@darmajaya.ac.id

Abstract – "The government's decision to discuss the RUU *Omnibus law* on health has become a controversial topic in society, especially among users of the Twitter social media platform. Users express their opinions regarding their stance on the RUU *Omnibus law* through tweets on Twitter. With diverse comments from users, it is essential to classify and visualize them into useful information about the positive and negative sentiments towards the RUU on health. This is crucial to understand the public's response to this policy. A total of 2406 sentiment data from Twitter users were collected using the RapidMiner software. Before analyzing the data using the K-Nearest Neighbors (KNN) algorithm, data preprocessing was carried out. After preprocessing, 2.406 data points were obtained, which were then divided into 1.684 tweets for testing and 722 tweets for training. The data was then processed using the KNN algorithm model executed in the RapidMiner software. The results of the data processing were presented in the form of tables, graphs, and word clouds, aligning with the research objective of providing clear and easily understandable visualizations about the RUU on health. This facilitates understanding for stakeholders without technical backgrounds to grasp the meaning and sentiments expressed. The research results indicate that the testing of K-Nearest Neighbors (KNN) yielded a high accuracy value, making it well-visualized at 84.58%. This indicates that the KNN model is highly successful in analyzing Twitter users' opinions on the Health Omnibus Law based on the data used and its ability to visualize effectively.

Keywords - data visualization, sentiment analysis, RUU *Omnibus law* on health, K-Nearest Neighbors (KNN), RapidMiner.

Abstrak – Keputusan pemerintah melakukan pembahasan terkait RUU *omnibus law* kesehatan menjadi perbincangan yang kontroversial di masyarakat, khususnya pengguna media sosial twitter. Pengguna memberikan komentarnya terkait sikap mereka terhadap RUU *omnibus law* melalui menu tweet pada twitter. Komentar yang beragam perlu di twitter penting untuk diklasifikasikan dan divisualisasikan agar menjadi informasi yang bermanfaat dari sentimen positif dan negatif terhadap RUU Kesehatan. Hal ini penting dilakukan supaya dapat diketahui bagaimana respon masyarakat terhadap kebijakan ini. Sebanyak 2406 data sentimen pengguna twitter dikumpulkan data dari twitter menggunakan software RapidMiner. Sebelum data analisis menggunakan algoritma KNN, dilakukan *preprocessing* data terlebih dahulu. Setelah dilakukan *preprocessing* diperoleh sebanyak 2.406 data yang kemudian dibagi menjadi data testing sebanyak 1.684 tweet dan data *training* sebanyak 722 tweet. Data kemudian diproses menggunakan model algoritma KNN yang dijalankan menggunakan software RapidMiner. Hasil dari pengolahan data disajikan visualisasi dalam bentuk tabel, grafik, maupun *word cloud*. Hal ini sesuai dengan tujuan penelitian ini yaitu untuk melakukan visualisasi data supaya dapat memberikan

gambaran yang jelas dan mudah dipahami tentang RUU Kesehatan sehingga memudahkan bagi pihak yang berkepentingan dan tidak memiliki latar belakang teknis untuk memahami makna dan sentimen tersebut. Hasil penelitian menunjukkan bahwa pengujian KNN menghasilkan nilai akurasi yang tinggi sehingga hasilnya dapat divisualisasikan dengan baik yaitu sebesar 84,58% yang mengindikasikan bahwa model KNN sangat berhasil dalam menganalisis pendapat pengguna twitter tentang RUU Omnibus Law Kesehatan berdasarkan data yang digunakan dan kemampuannya untuk memvisualisasikan dengan baik.

Kata Kunci – visualisasi data, analisa sentimen, RUU omnibuslaw Kesehatan, KNN, RapidMiner

I. PENDAHULUAN

DPR RI telah secara resmi mengesahkan RUU Kesehatan *Omnibus law* menjadi Undang-Undang melalui Rapat Paripurna DPR RI Masa Persidangan V tahun sidang 2022-2023 pada tanggal 11 Juli 2023 [1]. Meskipun disahkan, ada sejumlah pihak yang menyatakan ketidaksetujuan mereka terhadap pengesahan UU ini karena dianggap terlalu terburu-buru. Mereka berpendapat bahwa RUU Kesehatan ini baru saja dibahas pada tahun sebelumnya. Di DPR sendiri, Fraksi Partai Demokrat dan Fraksi PKS menjadi dua fraksi di DPR RI yang menolak menyetujui disahkannya Rancangan Undang-Undang (RUU) tentang Kesehatan [2].

Sebelum disahkan menjadi Undang-Undang, RUU Kesehatan menjadi perdebatan di masyarakat. Rencana ini banyak mendapat penolakan. Setidaknya terdapat lima organisasi profesi yang menolak, yaitu Ikatan Dokter Indonesia (IDI), Persatuan Dokter Gigi Indonesia (PDGI), Ikatan Apoteker Indonesia (IAI), Ikatan Bidan Indonesia (IBI), dan Persatuan Perawat Nasional Indonesia (PPNI) [3]. Selain terdapat pihak yang menolak, terdapat juga banyak masyarakat yang menyatakan dukungannya atas rencana tersebut [4]. Penolakan dan dukungan ini selain disampaikan secara langsung juga melalui media sosial.

Media sosial menjadi salah satu alat utama untuk berkomunikasi dan berbagi informasi dan mengemukakan pendapat atau opini mereka untuk berbagai topik [5]. Twitter merupakan salah satu platform media sosial yang paling populer digunakan oleh masyarakat untuk berbagi pendapat, pemikiran, dan komentar tentang suatu topik tertentu. Hal ini membuat Twitter menjadi sumber data yang berharga untuk melakukan analisis sentimen.

Twitter merupakan salah satu media sosial yang memberikan penggunanya mengirim dan membaca pesan yang berbasis teks [6]. Media sosial Twitter memungkinkan pengguna meninggalkan komentar dan pesan yang dapat

*) penulis korespondensi: Tupari

Email: tupari.2121210017@mail.darmajaya.ac.id

dibaca oleh sesama pengguna Twitter. Fitur untuk meninggalkan komentar dan pesan ini biasanya disebut sebagai Tweet atau Kicauan berupa pendapat atau komentar. Twitter memungkinkan pengguna membuat tweet hingga 140 karakter, menjadikan media sosial ini unik [7].

Begitu juga dengan isu aktual saat ini yang ramai dibahas di media sosial Twitter, pengguna menyampaikan komentarnya terkait RUU Kesehatan melalui menu tweet. Juga bermunculan hastag penolakan, salah satunya #TolakOmnibuslawKesehatan. Hal ini merupakan hal yang wajar karena penggunaan media sosial Twitter ini tidak hanya untuk berkomunikasi dan tetap terhubung dengan teman, keluarga, dan teman sekerja melalui pertukaran pesan yang cepat dan sering, tetapi juga digunakan untuk membahas masalah-masalah yang sedang berlangsung.

Beragamnya komentar dan pandangan masyarakat ini perlu untuk diklasifikasikan agar menjadi pengetahuan dan informasi baru yang bermanfaat. Pengetahuan ini berupa visualisasi data hasil analisis sentimen positif dan negatif terhadap RUU Kesehatan. Visualisasi data ini dapat memberikan gambaran yang jelas dan mudah dipahami tentang RUU Kesehatan. Informasi yang beragam dan kompleks dari analisis sentimen ini dapat diwakili dengan bentuk visual seperti grafik atau diagram dapat memudahkan bagi pihak yang berkepentingan dan tidak memiliki latar belakang teknis untuk memahami makna dan sentimen tersebut. Hal ini dimungkinkan dilakukan menggunakan analisis sentimen yang merupakan proses memahami, mengekstraksi, dan mengolah data teks secara otomatis untuk mendapatkan informasi emosional yang terkandung dalam kalimat.

Analisis sentimen yang disebut juga *opinion mining* adalah suatu bentuk pengetahuan yang diperoleh dari data *mining* yang ditujukan untuk menganalisis, memahami, mengolah, dan mengekstraksi data tekstual berupa opini tentang entitas seperti produk, organisasi, orang, atau topik tertentu [8]. Analisis sentimen adalah salah satu metode yang digunakan untuk mengidentifikasi opini atau perasaan yang diungkapkan dalam teks atau dokumen dan mengelompokkannya menjadi opini positif dan negatif. Pada dasarnya, metode ini bekerja dengan menganalisis teks untuk menentukan emosi mana yang positif atau negatif di dalamnya [9]. Hasil dari proses bisnis tersebut seseorang pelajari langkah selanjutnya untuk memperbaiki kelemahan dan kekuatan yang diketahui dari berbagai opini yang disajikan [7].

Permasalahan kurangnya pemahaman tentang makna sentimen pengguna twitter terhadap RUU *Omnibus law* Kesehatan dapat diatasi dengan penyajian informasi data melalui visualisasi dalam bentuk grafik maupun *word cloud*. Visualisasi data analisis sentimen merupakan alat komunikasi yang efektif untuk menyampaikan hasil penelitian kepada masyarakat. Informasi yang disajikan dalam bentuk visual yang menarik dapat meningkatkan pemahaman dan partisipasi publik dalam isu-isu yang berkaitan dengan RUU Kesehatan.

Dengan demikian, tujuan dari penelitian ini adalah untuk memberikan gambaran yang lebih jelas dan mudah dipahami

oleh masyarakat terkait sentimen pengguna twitter terhadap RUU *Omnibus law* Kesehatan. Penelitian ini juga bertujuan untuk memberikan informasi yang kompleks dari analisis sentimen dengan grafik atau diagram yang lebih intuitif, memudahkan pemahaman bagi pemangku kepentingan sehingga dapat diidentifikasi isu-isu tertentu yang menjadi fokus perhatian masyarakat. Hal ini dapat membantu pihak yang berkepentingan dalam menentukan prioritas penanganan dan komunikasi terhadap isu-isu yang paling signifikan atau kontroversial.

Visualisasi data dalam penelitian ini ditampilkan dalam bentuk *word cloud*, yaitu untuk menampilkan kata-kata yang paling sering muncul dalam sentimen positif, negatif, dan netral. Selain itu juga ditampilkan dalam bentuk *confusion matrix*, yaitu menampilkan performa model klasifikasi KNN melihat seberapa baik model tersebut dalam mengklasifikasikan sentimen.

Kemampuan untuk melihat tren sentimen dari waktu ke waktu memberikan wawasan tentang perubahan opini dan reaksi masyarakat terhadap RUU Kesehatan. Ini membantu dalam memahami bagaimana pandangan publik dapat berubah seiring berjalannya waktu.

Untuk melakukan penelitian ini, peneliti menggunakan software RapidMiner, yaitu perangkat lunak analisis data yang dapat digunakan untuk visualisasi data yang mudah. Di RapidMiner, dapat digunakan berbagai visualisasi untuk menggambarkan data seperti tabel, grafik, *word cloud*, diagram batang, histogram, dan lainnya.

II. PENELITIAN YANG TEKAIT

Beberapa penelitian sebelumnya telah dilakukan mengenai metode klasifikasi tersebut, antara lain Angelina Puput Giovanni et al., [10] yang menganalisis sentimen program Ruang Guru di Twitter menggunakan klasifikasi algoritma *Naive Bayesian* (NB). Algoritma, *Support Vector Machine* (SVM), *K-Nearest Neighbor* (K-NN) dan pemilihan fitur dengan algoritma PSO (*Particle Swarm Optimization*). Dari hasil penelitian disimpulkan bahwa algoritma PSO berbasis SVM merupakan aplikasi optimasi terbaik. Nilai akurasi yang diperoleh adalah 78,55 dan AUC 0,853.

Aprilia Wandani, Fauziah, Andrianingsih [7] menghasilkan penelitian bahwa algoritma K-NN dan *Random Forest* memiliki kinerja yang kurang baik dalam data sampel kecil. Hal ini berbeda dengan algoritma *Naive Bayes* yang memiliki akurasi lebih stabil pada data sampel besar atau kecil. Bitu Parga Zen, dkk [8] yang melakukan klasifikasi dengan algoritma *Support Vector Machine* (SVM) untuk memprediksi kelas berdasarkan model atau pola dari hasil proses training. Pada penelitian ini analisis sentimen dilakukan guna mengetahui respon masyarakat mengenai vaksin Sinovac dan didapatkan akurasi sebesar 74%.

Valentino Kevin Sitanayah Que, dkk [11] menggunakan RapidMiner untuk proses analisis analisis. Hasil analisis sentimen positif menggunakan SVM mencapai 62%, sedangkan sentimen negatif sebesar 38%. Sementara itu, pada SVM-PSO, opini positif sebesar 53% dan opini negatif sebesar 47%.

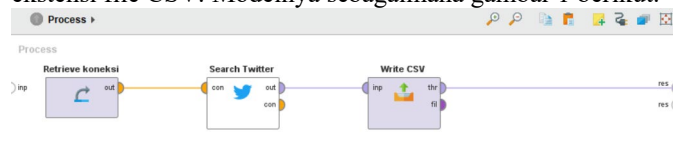
Guslendra dan Sarjon Defit [12], penelitian dengan judul *K-Means and K-NN Methods For Determining Student Interest* menyimpulkan bahwa hasil pengujian yang dilakukan menggunakan RapidMiner, hasil perhitungan dari metode-metode yang digunakan memiliki kemiripan. Ini menunjukkan bahwa hasil perhitungan dari kedua metode tersebut memiliki kesamaan atau kemiripan.

Dengan demikian, perbedaan penelitian ini dengan penelitian sebelumnya terletak pada fokus penelitian yaitu analisis sentimen terkait dengan RUU *Omnibus law* Kesehatan. Hal ini merupakan perbedaan utama dalam objek penelitian dibandingkan dengan penelitian sebelumnya yang memiliki fokus pada berbagai topik analisis sentimen yang berbeda, seperti program Ruang Guru di Twitter, respon masyarakat terhadap vaksin Sinovac, dan minat siswa. Perbedaan lain yaitu pada penelitian sebelumnya menggunakan berbagai perangkat lunak analisis, seperti RapidMiner dan algoritma PSO, sedangkan dalam penelitian ini juga menggunakan RapidMiner, tetapi fokus pada algoritma K-NN sehingga dapat dianggap sebagai perbedaan dalam pendekatan analisis. Kebaruan dalam penelitian ini dapat terletak pada fokus yang lebih khusus yaitu RUU *Omnibus law* Kesehatan, serta penggunaan metode K-NN untuk analisis sentimen, yang dapat memberikan wawasan yang berbeda dan berguna dalam konteks tersebut.

III. METODE PENELITIAN

A. Pengumpulan Data (*Crawling*)

Peneliti menggunakan software RapidMiner versi 9.10.013 untuk melakukan pengumpulan data yaitu komentar pengguna media sosial Twitter dengan kata kunci "*omnibus law* kesehatan" berbahasa Indonesia. Data *crawling* pada bulan Januari dan Mei 2023 dengan jumlah data bersih sebanyak 2406 tweet yang dibatasi hanya di wilayah Indonesia. Data kemudian disimpan ke komputer dengan ekstensi file CSV. Modelnya sebagaimana gambar 1 berikut.



Gbr. 1 Model *Crawling* data

B. Pengolahan Data

Selanjutnya dilakukan pengolahan data atau *Preprocessing* yang meliputi *case folding*, *tokenization*, *filterisasi*, *stopword removal*, dan *stemming*. *Case folding* yaitu mengubah seluruh teks menjadi huruf kecil sehingga tidak terdapat perbedaan antara huruf besar dan kecil. *Tokenization* adalah memecah teks menjadi token atau kata-kata yang masing-masing dapat dianggap sebagai sebuah unit. *Filterisasi* yaitu menghilangkan karakter-karakter yang tidak perlu seperti tanda baca, URL, tagar (#), atau mention (@) yang terdapat pada tweet. *Stopword removal* adalah menghilangkan kata-kata yang sering muncul namun tidak memiliki kontribusi dalam analisis seperti 'di', 'yang', 'ke', 'dari', dan lainnya. Tahap *stemming* adalah mengubah kata-

kata dalam bentuk bervariasi menjadi bentuk dasarnya. Contohnya mengubah kata 'kemajuan', 'memajukan', dan 'maju' menjadi 'maju'.

Setelah tahapan *preprocessing* dilakukan, data dapat dilabeli menggunakan metode TF-IDF (*term frequency-inverse document frequency*) untuk menentukan bobot kata yang digunakan dalam analisis sentimen. TF-IDF merupakan metode pembobotan kata yang menghitung frekuensi kemunculan suatu kata pada dokumen tertentu, dibagi dengan jumlah kata pada dokumen tersebut, kemudian dikalikan dengan algoritma invers dari jumlah dokumen yang mengandung kata tersebut dibagi dengan total jumlah dokumen.

Nilai TF-IDF tergantung pada frekuensi kemunculan kata dalam dokumen [13]. Kata-kata yang tersusun dalam masing-masing tweet akan diberikan bobot pada tahapan ini. Caranya dengan mengalikan nilai *Term Frequency (TF)* dengan *Inverse Document Frequency (IDF)* [5]. Pengukuran bobot dilakukan dengan mencari kemunculan masing-masing kata yang terdapat dalam tweet pada *feature list* positif, negatif, dan netral. Tweet tersebut akan digunakan sebagai data latih dan juga data uji.

Menurut Salton, nilai IDF didapatkan dengan persamaan sebagai berikut[8]:

$$IDF = \log \left(\frac{D}{dfi} \right) \quad (1)$$

Dimana:

D = jumlah dokumen

dfi = jumlah kemunculan term terhadap D.

Perhitungan bobot TF-IDF(W) untuk setiap dokumen terhadap kata kunci dengan rumus:

$$W_{ij} = tf_{ij} \log \left(\frac{D}{dfi} \right) \quad (2)$$

Setiap kata dihitung frekuensi kemunculannya di setiap tweet dalam korpus pada penelitian ini. Bobotnya dihitung dengan formula TF-IDF. Kata-kata yang sering muncul di banyak tweet namun tidak memberikan informasi signifikan dihilangkan dengan menggunakan *stopword removal*, sementara kata-kata yang muncul lebih jarang namun memberikan informasi yang lebih penting mendapatkan bobot yang lebih tinggi.

C. Analisis Sentimen

Pada tahap analisis sentimen, data diolah untuk melakukan labeling dan penambahan atribut guna menentukan nilai sentimen dari data, apakah itu negatif, positif, atau netral. Langkah-langkah dalam analisis sentimen ini sebagai berikut.

- Mencari data tweet yang termasuk sentimen negatif, positif, dan netral di setiap baris.
- Melakukan analisis sentimen dari keseluruhan dokumen sebagai negatif, positif, atau netral.
- Mengelompokkan semua atribut yang memiliki skor sentimen yang sama.
- Membuat visualisasi data hasil analisis sentimen untuk interaksi antar pengguna.

Digunakan rumus Skor = (jumlah kata positif - jumlah kata negatif) untuk menghitung skor sentimen. Hasil dari perhitungan skor tersebut akan menentukan kategori sentimen dari suatu kalimat, yaitu kelas positif jika skor > 0, kelas negatif jika skor < 0, dan kelas netral jika skor = 0.

D. Implementasi Algoritma

Algoritma yang digunakan dalam menyelesaikan masalah penelitian ini yaitu *K-Nearest Neighbors* (KNN) dan dengan bantuan software RapidMiner. KNN merupakan salah satu dari algoritma klasifikasi yang memiliki konsep melihat jarak tetangga terdekat dengan data baru [7]. KNN adalah salah satu algoritma paling populer yang termasuk dalam daftar 10 algoritma yang sederhana namun efektif untuk berbagai kasus pengenalan pola [14]. Algoritma *k-Nearest Neighbors* (k-NN) adalah *classifier* berbasis contoh dasar yang tidak membuat representasi deklaratif eksplisit dari kategori, melainkan bergantung pada label kelas yang dilampirkan pada dokumen pelatihan (misalnya, dokumen pengujian) [15].

KNN bekerja dengan cara mencari k titik data terdekat (*nearest neighbors*) dari sebuah titik data uji (*test data*) dalam ruang fitur. Setelah menemukan k tetangga terdekat, KNN akan melakukan mayoritas voting untuk menentukan label kelas dari titik data uji. KNN digunakan untuk memprediksi label kelas dari sebuah titik data uji berdasarkan mayoritas label kelas dari k tetangga terdekatnya. Dalam metode ini, parameter k pada algoritma KNN.

RapidMiner merupakan perangkat lunak terbuka untuk menganalisis *data mining*, *text mining*, dan analisis prediktif. RapidMiner menggunakan berbagai teknik deskriptif dan prediktif untuk memberikan wawasan kepada pengguna dan membantu pengguna membuat keputusan terbaik. RapidMiner memiliki berbagai operator *data mining*, termasuk operator *input*, *output*, *preprocessing* data, dan visualisasi yang dapat digunakan untuk melakukan operasi *data mining* dengan mudah dan cepat. Semua proses dalam penelitian ini dari proses awal *crawling* data hingga visualisasi dilakukan dengan RapidMiner. Validasi diuji menggunakan metode 10 *k-fold cross-validation*.

Proses analisis sentimen dilakukan menggunakan aplikasi RapidMiner dengan menerapkan metode KNN. Hal ini bertujuan untuk memperoleh visualisasi yang lebih baik dari hasil analisis sentimen. Model yang telah dibentuk akan dievaluasi menggunakan data uji. Langkah-langkah implementasi algoritma KNN pada penelitian ini sebagai berikut.

- Mempersiapkan data latih dan data uji. Data latih adalah data yang akan digunakan untuk melatih model KNN, sedangkan data uji adalah data yang akan digunakan untuk menguji performa model yang telah dilatih.
- Menentukan parameter-parameter yang akan diuji seperti jumlah tetangga terdekat (k), fungsi jarak, dan sebagainya.
- Menghitung nilai fitness untuk setiap partikel dalam populasi. Fitness dapat didefinisikan sebagai akurasi

model KNN dengan menggunakan parameter yang direpresentasikan oleh partikel tersebut.

- Melakukan Uji performa model KNN yang telah dilatih dengan menggunakan data uji dengan menghitung akurasi dan metrik evaluasi lainnya untuk mengevaluasi performa model.

E. Visualisasi

Setelah semua tahapan dilakukan, data dapat divisualisasikan menggunakan teknik *word cloud* untuk menampilkan kata-kata kunci yang paling sering muncul dalam data RUU Kesehatan di Twitter. Visualisasi *word cloud* ini membantu memberikan gambaran visual tentang topik-topik yang paling banyak dibicarakan dan memudahkan analisis tentang sentimen yang muncul. Visualisasi membantu pengguna menganalisis dan bernalar tentang informasi dan bukti untuk membuat informasi yang kompleks lebih mudah dipahami dan dimengerti [16].

Visualisasi data analisis sentimen memungkinkan pengamatan tren sentimen dari waktu ke waktu. Dengan melihat perubahan sentimen seiring berjalannya waktu, penelitian ini memberikan wawasan tentang bagaimana opini dan persepsi masyarakat terhadap RUU Kesehatan berkembang dari masa ke masa.

Word cloud dapat digunakan untuk memvisualisasikan kata-kata kunci yang paling sering muncul dalam data RUU Kesehatan. *Word cloud* dapat membantu memberikan gambaran visual tentang topik-topik yang paling banyak dibicarakan dan memudahkan analisis tentang sentimen yang muncul.

Dalam keseluruhan, analisis sentimen terhadap data RUU Kesehatan di Twitter dengan menggunakan tahapan *preprocessing* kalimat, *clustering*, dan visualisasi *word cloud* dapat memberikan pengetahuan yang penting tentang persepsi masyarakat dan membantu pemerintah dalam merumuskan kebijakan yang lebih tepat dan responsif.

IV. HASIL DAN PEMBAHASAN

Proses analisis sentimen pada penelitian ini menggunakan tools RapidMiner versi 9.10.013. Data yang digunakan adalah data tweet dari hasil kata kunci “RUU *omnibus law* kesehatan” menggunakan aplikasi RapidMiner dengan total data sebanyak 2470 tweet dari 1 Januari 2023 hingga Mei 2023. Data tweet yang telah dikumpulkan kemudian dilakukan pembersihan data (*preprocessing*). Setelah dilakukan *preprocessing* diperoleh data sebanyak 2.406 yang kemudian dibagi menjadi data testing 1684 tweet dan data *training* 722 tweet. Data *training* dibagi secara manual sesuai kelasnya, yaitu sentimen positif atau sentimen negatif yang dianggap dapat merepresentasikan *preprocessing*.

Data yang telah bersih kemudian dilakukan pelabelan menggunakan model yang telah dibuat dan juga divalidasi secara manual untuk memastikan kebenaran dan keabsahan data. Hasilnya sebagaimana dalam tabel 1 berikut.

TABEL I
HASIL PELABELAN

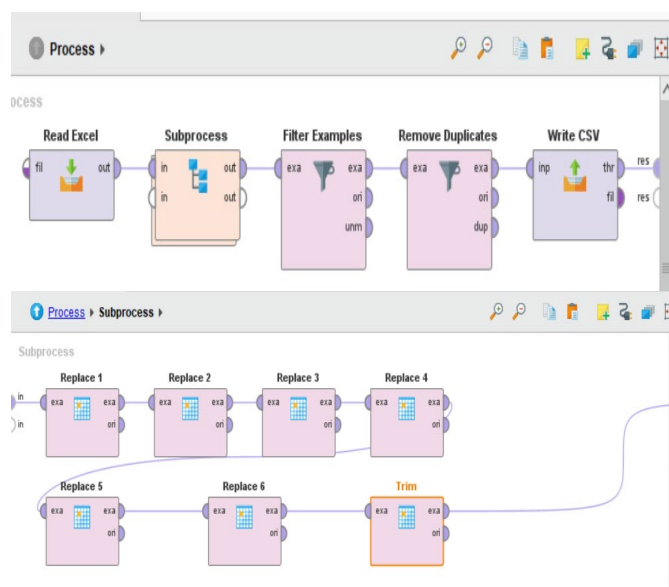
Label/Sentimen	Jumlah	Persentase
Positif	437	18%

Negatif	1925	80%
Netral	44	2%

Dari visualisasi pada tabel I di atas terlihat bahwa, hasil analisis sentimen RUU *Omnibus law* Kesehatan menggunakan metode *K-Nearest Neighbors (KNN)*, terdapat total 2,406 data yang telah dianalisis. Pada komentar positif dapat disimpulkan bahwa sebagian kecil dari data yang dianalisis memiliki sentimen positif terhadap RUU *Omnibus law* Kesehatan. Hal ini dimungkinkan terdapat aspek-aspek tertentu dalam RUU tersebut yang mendapat dukungan positif dari sebagian masyarakat. Pada komentar negatif, mayoritas data menunjukkan bahwa terdapat sentimen negatif terhadap RUU *Omnibus law* Kesehatan. Hal ini dimungkinkan menjadi petunjuk bahwa banyak orang merasa tidak puas atau khawatir terhadap isi RUU tersebut. Pada komentar netral, sentimen netral memiliki andil yang sangat kecil dalam dataset ini, sehingga mungkin tidak banyak informasi yang dapat diekstraksi dari kelompok ini.

Berdasarkan analisis sentimen menggunakan metode *K-Nearest Neighbors (KNN)*, mayoritas masyarakat cenderung memiliki sentimen negatif terhadap RUU *Omnibus law* Kesehatan. Penting untuk memahami alasan di balik sentimen negatif ini dan mempertimbangkan perbaikan atau klarifikasi yang mungkin diperlukan dalam RUU tersebut.

Model *preprocessing* data menggunakan software RapidMiner sebagai berikut.



Gbr. 2 Model *preprocessing*

Adapun hasilnya diperlihatkan pada tabel II sebagai berikut.

TABEL II
CONTOH HASIL PREPROCESSING

Tahap Preprocessing	Hasil
---------------------	-------

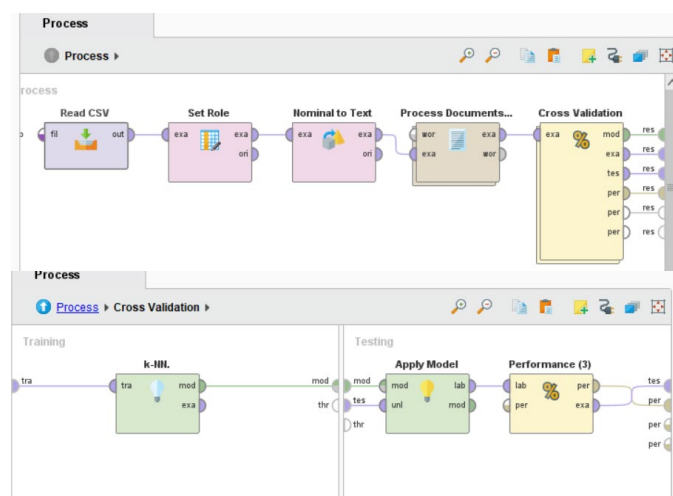
<i>Cleansing</i>	T1: Kajian terhadap RUU Kesehatan <i>Omnibus law</i> menjadi topik hangat untuk diperbincangkan di dunia maya T2: Salah satu alasan mengapa Pengadilan Ijazah Palsu tidak menunjukkan kemajuan berarti walaupun berpuluh Saksi sudah dihadirkan
<i>Case Folding</i>	T1: kajian terhadap ruu kesehatan <i>omnibus law</i> menjadi topik hangat untuk diperbincangkan di dunia maya T2: salah satu alasan mengapa pengadilan ijazah palsu tidak menunjukkan kemajuan berarti walaupun berpuluh saksi sudah dihadirkan
<i>Tokenizing</i>	T1: kajian; terhadap; ruu; kesehatan; omnibus; law; menjadi; topik; hangat; untuk; diperbincangkan; di; dunia; maya T2: salah; satu; alasan; mengapa; pengadilan; ijazah; palsu; tidak; menunjukkan; kemajuan; berarti; walaupun; berpuluh; saksi; sudah; dihadirkan;
<i>Normalization</i>	T1: kajian; terhadap; ruu; kesehatan; omnibus; law; menjadi; topik; hangat; untuk; diperbincangkan; di dunia; maya T2: salah; satu; alasan; mengapa; pengadilan; ijazah; palsu; tidak; menunjukkan; kemajuan; berarti; walaupun; berpuluh; saksi; sudah; dihadirkan
<i>Filtering</i>	T1: kajian; ruu; kesehatan; omnibus; law; topik; hangat; diperbincangkan; dunia; maya T2: alasan; pengadilan; ijazah; palsu; menunjukkan; kemajuan; walaupun; saksi; hadir
<i>Stemming</i>	T1: kajian; ruu; sehat; omnibus; law; topik; hangat; bincang; dunia; maya T2: alas; adil; ijazah; palsu; tunjuk; maju; walau; saksi; hadir

Pembobotan kata menurut TF-IDF dilakukan setelah dilakukan *preprocessing*. Bobot nilai pada masing-masing tweet yang dihasilkan dari hasil pengolahan data pada atribut kemudian dibandingkan dengan setiap probabilitasnya menurut pembobotan TF-IDF. Hasil dari probabilitas atribut positif dan negatif dibandingkan sehingga diketahui atribut yang memiliki probabilitas yang lebih besar. Jika probabilitas dari opini positif melebihi opini negatif, maka dokumen tersebut merupakan opini positif dan begitu juga sebaliknya. Contoh hasil pembobotan sebagaimana gambar 3 berikut.

Row No.	B	prediction(B)	confidence...	confidence...	confidence...	Text
1	Positif	Positif	0.200	0.600	0.200	34 idah menyampaikan bahwa Omnibus ...
2	Negatif	Positif	0.200	0.600	0.200	Penolakan semakin masif
3	Positif	Positif	0.200	0.600	0.200	Masyarakat juga kudu Paham bahwa BPJ...
4	Positif	Positif	0.200	0.600	0.200	Omnibus Law Cipta Kerja pengambilan ...
5	Positif	Positif	0	0.875	0.125	Partai Buruh dan Organisasi Sentral Buru...
6	Positif	Positif	0.200	0.600	0.200	BPJS itu jg untu jaminan kematian, JHT, k...
7	Positif	Positif	0.200	0.600	0.200	Finally di era omnibus law uu cipta kerja...
8	Positif	Positif	0.200	0.600	0.200	Sipkosee itu bangun tol dr ngutang dan...
9	Positif	Positif	0.200	0.600	0.200	DPR TERGABUT, OMNIBUS DISAYANG, H...
10	Positif	Positif	0.200	0.600	0.200	On the other hand, Ketua BEM kalau mau ...
11	Positif	Positif	0.200	0.600	0.200	Apakah lahinya UU Kesehatan omnibus I...
12	Positif	Positif	0.200	0.600	0.200	Pengusaha meminta pemerintah tidak me...
13	Positif	Positif	0.200	0.600	0.200	sekarang...
14	Positif	Positif	0.200	0.600	0.200	APINDO memberikan respons setelah da...
15	Positif	Positif	0.200	0.600	0.200	Jelang Aksi Nasional Tolak Pempu Cipta ...

Gbr. 3 Contoh Hasil Pembobotan

Model untuk melakukan klasifikasi KNN menggunakan 10 *k-fold validation* pada *cross validation* sebagaimana gambar berikut ini.



Gbr. 4 Model Confusion Matrix Untuk KNN dengan 10 K-Fold

Pengujian metode klasifikasi KNN menggunakan 10 *k-fold validation* atau nilai *fold default* pada *cross validation*, yaitu dengan acak membagi data menjadi sepuluh bagian data dan prosesnya berulang sebanyak jumlah kelompok yang telah ditentukan. Adapun visualisasi nilai akurasi disajikan pada gambar 5 berikut ini.

	true Negatif	true Positif	true Neutral	class precision
pred. Negatif	1821	225	35	87.51%
pred. Positif	99	212	7	66.67%
pred. Neutral	5	0	2	28.57%
class recall	94.60%	48.51%	4.55%	

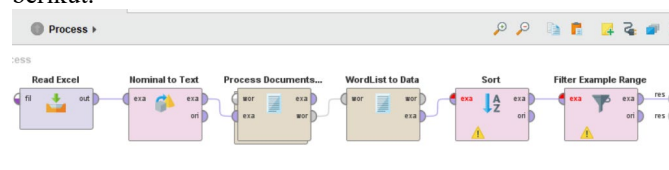
Gbr. 5 Hasil Confusion Matrix Untuk KNN dengan 10 K-Fold

Berdasarkan visualisasi melalui gambar di atas, terlihat nilai akurasi KNN sebesar 84,58% hal ini berarti model dapat memprediksi sentimen data dengan sangat baik. Akurasi

adalah metrik yang mengukur sejauh mana model dapat memprediksi label emosional dengan benar berdasarkan data uji yang digunakan dalam evaluasi.

Nilai akurasi sebesar 84,58% berarti sekitar 84,58% dari seluruh data eksperimen yang digunakan diprediksi dengan benar oleh model. Hal ini menunjukkan bahwa model KNN sangat efektif dalam melakukan analisis opini *RUU omnibus law* kesehatan berdasarkan data yang digunakan untuk melatih dan memvisualisasikannya dengan baik.

Visualisasi hasil analisis sentimen menggunakan *word cloud* pada RapidMiner dilakukan untuk melihat gambaran frekuensi kata-kata yang sering muncul. Tujuan visualisasi ini untuk memudahkan pemahaman dan interpretasi hasil analisis sentimen secara visual. Adapun model yang disusun sebagai berikut.



Gbr. 6 Model visualisasi data dengan Word cloud pada RapidMiner

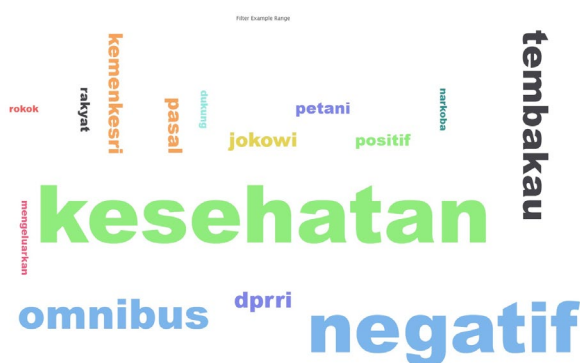
Setelah model disusun, kemudian dilanjutkan dengan proses pengujian yang menghasilkan output sebagai berikut.

Row No.	word	in documents	total	in class (No.)	in class (Po.)	in class (Net.)	in class (437)	in class (19..)	in class (44)	in clas
1	kesehatan	2128	3264	2535	671	58	0	0	0	0
2	tembakau	956	1452	1427	17	8	0	0	0	0
3	omnibus	1119	1234	886	226	22	0	0	0	0
4	pasal	552	651	605	45	1	0	0	0	0
5	dprn	624	626	607	13	6	0	0	0	0
6	jokowi	543	550	533	17	0	0	0	0	0
7	kemenkesri	539	539	518	20	1	0	0	0	0
8	petani	434	470	465	4	1	0	0	0	0
9	raliyat	356	439	383	39	7	0	0	0	0
10	naraboa	302	322	318	3	1	0	0	0	0
11	rokok	283	316	310	6	0	0	0	0	0
12	dukung	306	314	304	8	2	0	0	0	0
13	mengalahkan	278	279	279	0	0	0	0	0	0
14	sistem	171	263	191	70	2	0	0	0	0
15	pelawaja	233	262	258	4	0	0	0	0	0

Gbr. 7 Filter 15 Kata yang paling sering muncul

Berdasarkan tabel di atas terlihat bahwa kata "Kesehatan" mendominasi *word cloud*, yang mengindikasikan bahwa isu kesehatan adalah salah satu aspek paling penting dalam RUU ini. Sentimen terhadap isu ini sangat beragam dan tergantung pada pandangan individu dan kelompok. Kata "Tembakau" juga muncul cukup besar. Hal ini menunjukkan bahwa regulasi terhadap produk tembakau merupakan bagian penting dari RUU ini dan menjadi tweet yang cukup mendominasi di twitter. Sentimen terhadap regulasi tembakau ini juga bervariasi tergantung pada perspektif individu dan industri tembakau itu sendiri. Penggunaan kata "Omnibus" dalam *word cloud* menunjukkan bahwa RUU ini dirancang untuk merangkul berbagai isu yang beragam dalam satu undang-undang. Sentimen pengguna twitter terhadap pendekatan ini beragam, ada yang menganggapnya efisien sementara yang lain menolak dengan berbagai alasan. Jika divisualisasikan

dengan tampilan *word cloud* pada RapidMiner sebagaimana gambar 8 dan gambar 9 berikut.



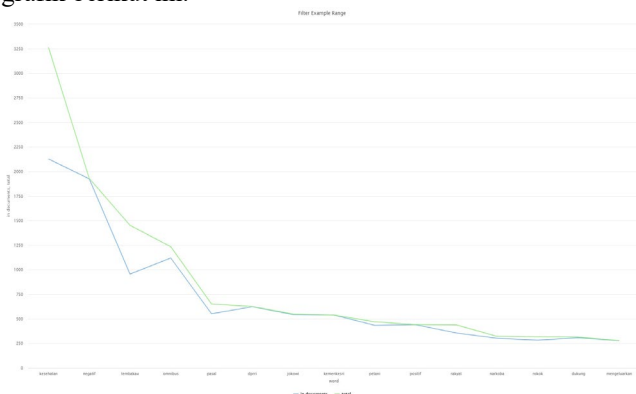
Gbr. 8 Visualisasi *Word Cloud* omnibus law Kesehatan dengan RapidMiner (in document)

Sedangkan jika divisualisasikan dari total kata diperoleh gambaran sebagai berikut.



Gbr. 9 Visualisasi *Word Cloud* omnibus law Kesehatan dengan RapidMiner (in total document)

Perbandingan dua gambar tersebut, dapat ditampilkan dalam grafik berikut ini.



Gbr. 10 Grafik perbandingan word cloud kata yang sering muncul in document dan in total document

Berdasarkan gambar di atas, *word cloud* menampilkan kata dengan frekuensi terbanyak, semakin banyak frekuensi kata tersebut semakin menonjol dibanding kata yang lain. Beberapa kata dengan frekuensi terbanyak yaitu Kesehatan, tembakau, dan omnibus. Frekuensi ini menggambarkan

respon pengguna twitter memberi respon positif, negatif, atau netral terhadap RUU *omnibus law* kesehatan.

V. KESIMPULAN

Melalui penggunaan algoritma KNN, penelitian ini berhasil mengidentifikasi sentimen positif dan negatif dari masyarakat terkait RUU *Omnibus law* Kesehatan, termasuk juga melakukan visualisasi data. Melalui visualisasi data, informasi ini memberikan gambaran tentang pandangan dan reaksi masyarakat terhadap undang-undang tersebut. Hasil algoritma KNN menunjukkan efektivitas penggunaan KNN dalam meningkatkan kualitas analisis sentimen dengan nilai akurasi 84,58% yang mengindikasikan KNN bekerja dengan sangat baik dalam menganalisis sentimen pada konteks penelitian ini.

Kelebihan dari penelitian ini visualisasi data analisis sentimen memberikan gambaran yang jelas dan informatif tentang pandangan masyarakat terhadap RUU *Omnibus law* Kesehatan. Grafik dan diagram yang menarik membantu dalam memahami pola sentimen dengan lebih mudah.

Kekurangan dari penelitian ini adalah keterbatasan algoritma tanpa menggunakan algoritma lain untuk membantu meningkatkan akurasi, meskipun hasil akurasi sudah sangat tinggi. Penggunaan algoritma klasifikasi yang lebih kompleks atau teknik *deep learning* dapat menjadi opsi untuk meningkatkan performa analisis sentimen. Menurut Dwitama [17] algoritma *Recurrent Neural Network* (RNN) mempunyai hasil akurasi terbaik dibandingkan enam (6) algoritma lain.

Kemungkinan pengembangan penelitian selanjutnya dapat menggabungkan data dari berbagai sumber lain seperti media sosial lainnya, berita, atau survei publik sehingga dapat memberikan perspektif lain dan komprehensif tentang pandangan masyarakat terhadap RUU Kesehatan. Penerapan teknik *deep learning* seperti penggunaan *neural networks* juga dapat membantu dalam meningkatkan akurasi analisis sentimen dan kemampuan untuk memproses teks dengan lebih kompleks.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Dosen pengampu mata kuliah Data Visualization, ibu Dr. Chairani, S.Kom., M.Eng dari Institut Informatika dan Bisnis Darmajaya, khususnya program Magister Teknik Informatika atas dukungannya dalam penelitian ini.

DAFTAR PUSTAKA

- [1] DPR-RI, "Paripurna DPR Tetapkan UU Kesehatan, Regulasi Komprehensif di Bidang Kesehatan," 2023. [https://www.dpr.go.id/berita/detail/id/45462/t/Paripurna DPR Tetapkan UU Kesehatan, Regulasi Komprehensif di Bidang Kesehatan#:~:text=Rapat Paripurna DPR RI mengesahkan,11%2F7%2F2023](https://www.dpr.go.id/berita/detail/id/45462/t/Paripurna%20DPR%20Tetapkan%20UU%20Kesehatan,%20Regulasi%20Komprehensif%20di%20Bidang%20Kesehatan#:~:text=Rapat%20Paripurna%20DPR%20RI%20mengesahkan,11%2F7%2F2023). (diakses Sep 17, 2023).
- [2] DPR-RI, "Dua Fraksi di DPR Tolak Pengesahan UU Kesehatan," 2023. [https://www.dpr.go.id/berita/detail/id/45459/t/Dua Fraksi di DPR Tolak Pengesahan UU Kesehatan](https://www.dpr.go.id/berita/detail/id/45459/t/Dua%20Fraksi%20di%20DPR%20Tolak%20Pengesahan%20UU%20Kesehatan) (diakses Sep 17, 2023).
- [3] Z. Abidin, "RUU 'Omnibus law' Kesehatan Versus Uniknya Profesi Dokter," 2023. <https://www.idionline.org/article/ruu-omnibus-law-kesehatan-versus-uniknya-profesi-dokter> (diakses Sep 17, 2023).

- [4] N. KATINGKA, "Perlu Transformasi, Koalisi Organisasi Nakes Dukung RUU Kesehatan," 2023. <https://www.kompas.id/baca/humaniora/2023/05/06/organisasi-nakes-dukung-transformasi-lewat-ruu-kesehatan> (diakses Sep 17, 2023).
- [5] Z. Alhaq, A. Mustopa, S. Mulyatun, dan J. D. Santoso, "Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter," *J. Inf. Syst. Manag.*, vol. 3, no. 2, hal. 44–49, 2021, doi: 10.24076/joism.2021v3i2.558.
- [6] H. Sujadi, "Analisis Sentimen Pengguna Media Sosial Twitter Terhadap Wabah Covid-19 Dengan Metode Naive Bayes Classifier Dan Support Vector Machine," *Infotech J.*, vol. 8, no. 1, hal. 22–27, 2022, doi: 10.31949/infotech.v8i1.1883.
- [7] A. Wandani, Fauziah, dan Andrianingsih, "Sentimen Analisis Pengguna Twitter pada Event Flash Sale Menggunakan Algoritma K-NN, Random Forest, dan Naive Bayes," *J-Sakti*, vol. 5, no. 2, hal. 651–665, 2021.
- [8] B. P. Zen, D. Wicaksana, dan H. Alfidzar, "Analisis Sentimen Tweet Vaksin Covid 19 Sinovac Menggunakan Metode Support Vector Machine," *Jdmsi*, vol. 3, no. 2, hal. 21–27, 2022.
- [9] F. Hashfi, D. Sugiarto, dan I. Mardianto, "Sentiment Analysis of An Internet Provider Company Based on Twitter Using Support Vector Machine and Naive Bayes Method," *J. Tek. Inform.*, vol. 14, no. 1, hal. 1–6, 2022.
- [10] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, dan W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *J. Teknoinfo*, vol. 14, no. 2, hal. 115, 2020, doi: 10.33365/jti.v14i2.679.
- [11] V. Kevin, S. Que, A. Iriani, dan H. D. Purnomo, "Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization (Online Transportation Sentiment Analysis Using Support Vector Machine Based on Particle Swarm Optimization)," vol. 9, no. 2, hal. 162–170, 2020.
- [12] G. Guslendra, S. Defit, dan R. Bastola, "K-Means and K-NN Methods For Determining Student Interest," *Int. J. Artif. Intell. Res.*, vol. 6, no. 1, 2022, doi: 10.29099/ijair.v6i1.222.
- [13] V. Dwi Antonio, S. Efendi, dan H. Mawengkang, "Sentiment analysis for covid-19 in Indonesia on Twitter with TF-IDF featured extraction and stochastic gradient descent," *Int. J. Nonlinear Anal. Appl.*, vol. 13, no. 1, hal. 2008–6822, 2022, [Daring]. Tersedia pada: <http://dx.doi.org/10.22075/ijnaa.2021.5735>.
- [14] Y. Setiawan, "Data Mining berbasis Nearest Neighbor dan Seleksi Fitur untuk Deteksi Kanker Payudara," *J. Pengemb. IT*, vol. 8, no. 2, hal. 89–96, 2023, [Daring]. Tersedia pada: <http://ejournal.poltektegal.ac.id/index.php/informatika/article/view/4994>.
- [15] S. Sahara dan R. A. Permana, "Metode KNN Pada Sentiment Analisis Review Produk Game Android," *Indones. J. Netw. Secur.*, vol. 11, no. 2, hal. 123–128, 2022.
- [16] E. Undamayanti, T. I. Hermanto, dan I. Kaniawulan, "Analisis Sentimen Menggunakan Metode Naive Bayes Berbasis Particle Swarm Optimization Terhadap Pelaksanaan Program Merdeka Belajar Kampus Merdeka," *J. Sains Komput. Inform.*, vol. 6, no. 2, hal. 916–930, 2022.
- [17] A. P. J. Dwitama, "Deteksi Ujaran Kebencian Pada Twitter Bahasa Indonesia Menggunakan Machine Learning: Reviu Literatur," *J. SNATI*, vol. 1, no. 1, hal. 31–39, 2021.