

# ANALISIS SENTIMEN RUU KUHP DI MEDSOS TWITTER DENGAN METODE NAÏVE BAYES

Husni Muzaki, Hendra Marcos

*Program Studi Informatika, Universitas Amikom Purwokerto*

Jl. Letjend Pol. Soemarto No.127, Watumas, Purwanegara, Kec. Purwokerto Utara, Kabupaten Banyumas, Jawa Tengah 53127

khusnimuzaki018@gmail.com, hendra.marcos@amikompurwokerto.ac.id

**Abstract** - Sentiment analysis is the process of making decisions on opinions or comments, whether positive, negative, or neutral by using text analysis from various internet data sources. The internet makes social media the basic basis for building technology, including social media Twitter. Twitter has become very popular with its users because the news information it conveys is relatively accurate and real time. This research tries to raise a hot theme regarding the ratification of the pros and cons of RUU KUHP by the DPR RI. The Naïve Bayes algorithm is used in this sentiment analysis by means of document-based data mining (text mining). Based on previous studies, the use of classification models and accuracy with the Naïve Bayes method yields very good results, even if it uses a lot or a little data. Opinion data collection is done by crawling data on Twitter in an online document. The software used for this sentiment opinion analysis is RapidMiner Studio 10.0. The software is very easy to use because it is open-source based and can analyze data mining and predict accuracy automatically and online. The results of this study indicate that the level of accuracy for opinion analysis on the Criminal Code Bill is 100.00%, getting positive and negative opinions from class precision and recall each 100.00%. By looking at the results of its accuracy, the Naïve Bayes method for this study to get maximum results.

**Keywords** - Sentiment Analysis, RUU KUHP, Twitter, Naïve bayes.

**Abstrak** - Analisis sentimen merupakan proses pengambilan keputusan opini atau komentar, baik positif, negatif, atau netral dengan menggunakan analisis teks dari berbagai sumber data internet. Internet menjadikan media sosial sebagai basis dasar dalam membangun teknologi, termasuk media sosial Twitter. Twitter menjadi sangat digemari penggunaannya karena informasi berita yang disampaikan relatif akurat dan real time. Penelitian mencoba untuk mengangkat tema yang sedang hangat mengenai pro-kontra pengesahan RUU KUHP oleh DPR RI. Algoritma Naïve Bayes digunakan di analisis sentimen ini dengan cara menambang data (text mining) berbasis dokumen. Berdasarkan penelitian-penelitian sebelumnya, penggunaan model klasifikasi dan akurasi dengan metode Naïve Bayes mendapatkan hasil yang sangat baik, walaupun menggunakan banyak atau sedikit data. Pengambilan data opini dilakukan dengan cara crawling data di Twitter pada sebuah dokumen secara online. Software yang digunakan untuk analisis sentimen opini ini yaitu RapidMiner Studio 10.0. Software tersebut sangat mudah digunakan karena berbasis open-source dan dapat menganalisa data mining dan akurasi prediksi secara otomatis dan online. Hasil penelitian ini menunjukkan bahwa tingkat akurasi untuk analisis opini RUU KUHP yaitu 100.00% , mendapatkan opini positif dan negatif dari class precision dan recall masing-masing 100.00%. Dengan melihat hasil akurasinya, metode Naïve Bayes untuk penelitian ini mendapatkan hasil yang maksimal.

**Kata Kunci** - Analisis Sentimen, RUU KUHP, Twitter, Naïve Bayes

## I. PENDAHULUAN

Di era digital ini, perkembangan jejaring sosial banyak digunakan sebagai sarana komunikasi populer, salah satunya adalah jejaring sosial *Twitter* yang digunakan untuk iklan, produk, kampanye politik dan sebagai sarana menyampaikan pendapat. kritik , proposal dan solusi publik [1]. Menurut [2] dalam mencari informasi yang akurat, *twitter* dapat digunakan sebagai media sosial favorit. Namun, jika digunakan, pengguna bertanggung jawab untuk menganalisis informasi dengan benar untuk mendukung keputusan atau pilihan pihak tersebut.

Dengan didapatkannya informasi yang akurat melalui analisis, Pengguna *Twitter* memanfaatkannya untuk mendapatkan informasi berupa kritik dan rekomendasi terpercaya, misalnya pengesahan KUHP DPR RI. Semakin banyak pendapat atau pengaduan dari orang yang lebih banyak dapat membentuk pendapat yang kuat dan dapat digunakan sebagai kontribusi positif terhadap hasil dan keputusan dari pihak terkait [3]. Analisis sentimen sebagai salah satu *machine learning* digunakan untuk analisis informasi dari opini *tweet*. Menurut [4] analisis sentimen adalah pengklasifikasian polaritas opini pada tingkat dokumen teks, kalimat, atau sudut pandang dengan menentukan apakah opini

yang diberikan dalam data dokumen itu positif, negatif, atau netral. Analisis sentimen juga dapat disebut sebagai proses pemahaman, penggalian, dan pengolahan informasi secara otomatis untuk memperoleh informasi berupa hasil opini yang tertuang dalam suatu pernyataan opini. Penelitian ini melakukan analisis opini untuk mengetahui seberapa besar kecenderungan pendapat masyarakat terhadap pengesahan RUU KUHP yang menjadi undang-undang yang mengandung perasaan negatif, positif atau netral.

Penjelasan umum tentang RUU KUHP tergambar dari kenyataan bahwa gagasan pembentukan RUU KUHP yang memasukkan hukum pidana masyarakat ke dalam perundang-undangan formal dimulai dari keseimbangan antara kepentingan atau perlindungan pribadi dan kepentingan atau perlindungan kelompok masyarakat antara kepastian hukum dan keadilan [5]. DPR RI mengesahkan Proyek Hukum Pidana (RUU KUHP) pada Selasa (12/6/2022). Dengan disahkannya RUU menjadi UU terdapat banyak sekali pro-kontra baik oleh masyarakat, mahasiswa, maupun pakar hukum di berbagai universitas di Indonesia. Pengesahan tersebut dianggap berpotensi terhadap pelanggaran HAM. Tujuan dari analisis sentimen ini adalah untuk menganalisa bagaimana opini publik yang sentimental terkait dengan isu ini melalui *Twitter*. Algoritma atau metode yang digunakan pada analisis opini ini adalah *Naive Bayes*.

*Naive Bayesian Classifier* (NBC) adalah metode *machine learning* untuk menganalisis sentimen. NBC adalah teknik dalam pembelajaran mesin yang berbasis probabilitas. NBC adalah metode klasifikasi teks sederhana, tetapi memiliki akurasi dan efisiensi performa yang sangat tinggi [6]. Masalah yang dibahas adalah bagaimana algoritma klasifikasi *Naive Bayes* dapat secara otomatis mengklasifikasikan tweet ke dalam kategori sentimen negatif, positif, atau netral dan mendapatkan nilai presisi dan recall, serta bagaimana algoritma klasifikasi *Naive Bayes* yang akurat dapat mengklasifikasikan teks menjadi sentimen negatif dan positif. Dalam penelitian menggunakan algoritma NBC, peneliti menggunakan software open source yaitu *RapidMiner*, yang dapat melakukan pengambilan data secara online.

## II. METODE PENELITIAN

Penelitian ini disusun sebagai proses penelitian induktif dalam mencari dan mengumpulkan data yang ada di media sosial *twitter* dan penelitian-penelitian sebelumnya dengan tujuan mendapatkan data yang dapat mendukung dalam menentukan hasil akurasi dari metode *naive bayes*.

### 1. Sumber Data

Penelitian ini dimulai dari mengumpulkan data, yang terdiri dari penelitian perpustakaan (*library research*), penelitian lapangan (*field research*), dan penelitian laboratorium (*laboratory research*) Media sosial *twitter* menjadi sumber data dalam penelitian ini. Salah satu jejaring sosial yang memberikan cuplikan informasi terkait berita-berita yang sedang hangat adalah *Twitter*. Seorang pengguna *twitter* dapat mengetahui informasi kasus beserta dengan akun yang membuat cuitannya. Informasi yang diberikan oleh *twitter* relatif cepat dalam update terhadap berita dan jarang mengalami down server. dikarenakan inti masalah yang ingin peneliti lakukan yaitu tentang sentimen pengesahan RUU KUHP, tentunya media sosial *twitter* termasuk yang tepat sebagai sumber data dikarenakan sebab-sebab seperti keterangan diatas.

### 2. Pengumpulan Data

Setelah mendapatkan data dari sumber media sosial *twitter*, langkah selanjutnya yaitu mengambil data berupa opini masyarakat tentang pengesahan RUU KUHP menjadi UU. Pengumpulan data dilakukan dengan *web crawling* data dengan mengambil data yang tersedia di *twitter* secara online untuk umum. Proses ini berlangsung dengan cara memasukkan data berupa informasi yang telah ditemukan ke dalam file disk komputer.

### 3. Study Literatur

Study literatur digunakan sebagai sumber penelitian untuk mencari rujukan atau referensi. Studi literatur bisa didapatkan dari berbagai literatur seperti jurnal, buku, karya tulis/makalah penelitian. Penelitian ini menggunakan teori/materi yang relevan dan saling berkesinambungan dengan query atau kata kunci seperti, sentimen analisis, *naive bayes*, *data mining*, *RapidMiner*, dan sebagainya. Inti dari pengambilan teori yang bersangkutan adalah menyelesaikan problem-problem dari penelitian yang sedang dilakukan.

### 4. Pre-processing

Data yang diperoleh dari hasil pengambilan data tidak dapat diklasifikasikan otomatis karena masih terdapat banyak simbol dan kata yang tidak diperlukan lagi pada data, sehingga diperlukan langkah *preprocessing* agar data lebih terstruktur dan terorganisir sehingga dapat diklasifikasikan. Tujuan dari *preprocessing* adalah untuk menyajikan dokumen sebagai vektor fitur, yaitu memisahkan teks menjadi kata tersendiri. Dalam pengklasifikasian yang

telah diusulkan, teks dokumen terlebih dahulu dimodelkan sebagai suatu peristiwa berupa transaksi. Pemilihan kata kunci, yang merupakan proses pemilihan fitur adalah langkah pra-pemrosesan terpenting pada proses indek dokumen. [7]

### III. HASIL DAN PEMBAHASAN

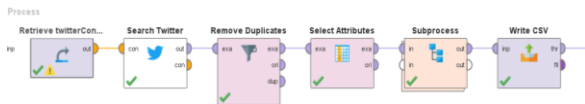
Dari Software yang dipakai pada proses analisis sentiment yaitu *RapidMiner Application*. *RapidMiner* menjadi salah satu software open source dan dilakukan secara online untuk tersambung ke akun twitter. Berikut tahapan dalam proses sentimen analisis sampai pada hasil akurasi.

#### A. Pengumpulan Data

Pengumpulan data diambil dari media sosial *Twitter*. Setelah mendapatkan data yang diinginkan, langkah selanjutnya adalah mengindeks/crawling informasi berupa data opini berupa tweet dari pengguna *Twitter*. Jumlah informasi data yang diambil adalah 235 ulasan, dimana kami memproses data sebagai ulasan negatif, positif, dan netral.



Gambar 1. Proses Pengumpulan Data



Gambar 2. Proses Crawling Data Twitter

Keterangan :

- *Retrieve Twitter Connection* : mengkoneksikan akun Twitter dengan *rapidminer*.
- *Search Twitter* : Mencari data melalui crawling.
- *Remove Duplikat* : Menghilangkan baris duplikat yang berulang.
- *Select Atribut* : Memilih satu atau lebih catatan yang berisi informasi spesifik.
- *Subprocess* : Menguji nilai true atau false pada nilai teks.
- *Write CSV* : Menulis data hasil crawling ke dalam format csv.

#### B. Pre-processing Data

Setelah Tahap *pre-processing* dapat dilihat pada Gambar 3.



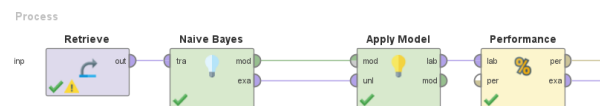
Gambar 3. Tahap Pre-processing

Keterangan :

- *Tokenize* : Menghapus karakter seperti tanda baca dan filter.
- *Transform Cases* : Mengubah huruf pada teks menjadi huruf kecil atau kapital.
- *Filter Stopword* : Sebelum mengkategorikan teks, hilangkan terlebih dahulu teks yang tidak padu untuk memperkecil dimensi teks.
- *Filter Token* : Menyaring Kembali karakter tertentu yang masih ada.
- *Stemming* : mengubah kata bersifat majemuk menjadi kata dasar.

#### C. Pengolahan Data

Dalam Tahap pengolahan data dapat dilihat pada Gambar 4.



Gambar 4. Tahap Cross Validation Naive Bayes

Tahap diatas merupakan akhir dari proses sentimen yaitu *Cross Validation*. Proses dimulai dari mengambil data yang sudah diambil dan di lakukan proses *preprocessing*. Kemudian, masuk ke dalam tahap memasukkan data ke dalam metode. Setelah data dimasukkan, dilakukan pengujian dengan proses training dan testing. Metode *Naive Bayes* akan menghasilkan nilai berupa akurasi, presisi, dan recall yang berdasarkan performa algoritma.

accuracy: 100.00%				
	true neutral	true positive	true negative	class precision
pred. neutral	203	0	0	100.00%
pred. positive	0	7	0	100.00%
pred. negative	0	0	1	100.00%
class recall	100.00%	100.00%	100.00%	

Gambar 5. Nilai Akurasi Naive Bayes

Dari hasil diatas kita ketahui bahwa nilai akurasi dari algoritma *Naive Bayes* adalah 100%, dengan nilai masing-masing kelas presisi dan recall adalah 100%.

### IV. KESIMPULAN

Berdasarkan analisis dan pengujian yang dilakukan dalam penelitian ini terhadap penerapan algoritma *naive bayes* pada RUU KUHP dapat ditarik kesimpulan bahwa :

1. Data menunjukkan bahwa akurasi pengesahan KUHP adalah 100,00%.

2. Hasil klasifikasi pada sentimen netral sangat mendominasi ketimbang sentimen positif dan negatif dengan 203 ulasan, 7 ulasan positif dan 1 ulasan negatif
3. Dapat disimpulkan bahwa lebih banyak orang yang ber opini netral terhadap pengesahan RUU KUHP dibandingkan dengan yang positif dan negatif.

#### DAFTAR PUSTAKA

- [1] A. V. Sudiantoro and E. Zuliarso, "Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma Naïve Bayes Classifier Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma Naïve Bayes Classifier," vol. 10, no. 2, pp. 69–73, 2018.
- [2] Y. S. Mahardhika and E. Zuliarso, *Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritma Naives Bayes Classifier*. 2018.
- [3] I. F. Rozi, E. N. Hamdana, M. Balya, and I. Alfahmi, "Pengembangan Aplikasi Analisis Sentimen Twitter Menggunakan Metode Naïve Bayes Classifier (Studi Kasus Samsat Kota Malang)." [Online]. Available: <https://path.com/p/3pB4Qs>
- [4] D. Ramadhan and E. B. Setiawan, "Analisis Sentimen Program Acara Di Sctv Pada Twitter Menggunakan Metode Naive Bayes Dan Support Vector Machine," ... *Telkomuniversity.Ac.Id*, vol. 6, no. 2, pp. 9736–9743, 2019, [Online]. Available: <https://openlibrarypublications.telkomuniversit y.ac.id/index.php/engineering/article/view/10708>
- [5] L. S. Widayati, "Perluasan Asas Legalitas Dalam Ruu Kuhp," *Negara Huk.*, vol. 2, no. 2, pp. 307–328, 2011.
- [6] D. M. E. D. M. Hussein, "A survey on sentiment analysis challenges," *J. King Saud Univ. - Eng. Sci.*, vol. 30, no. 4, pp. 330–338, 2018, doi: 10.1016/j.jksues.2016.04.002.
- [7] V. Srividhya and R. Anitha, "Evaluating Preprocessing Techniques in Text Categorization," *Int. J. Comput. Sci. Appl.*, vol. 47, no. 11, pp. 49–51, 2010.