**Analisis Sentimen Mengenai Undang - Undang TPKS pada Media Sosial Twitter Menggunakan Metode Support Vector Machine dan K-Nearest Neighbour**

**Arviandri Naufal Zaki1, Dian Pratiwi, ST, MTI2, Syandra Sari, S.Kom, M.Kom3**

1,2,3Jurusan Teknik Informatika, Fakultas Teknologi Industri, Universitas Trisakti

Jl. Letjen S. Parman No.1, RT.6/RW.16, Grogol, Kec. Grogol petamburan, Kota Jakarta Barat, Daerah Khusus Ibukota Jakarta 11440

[arviandri064001800035@std.trisakti.ac.id1](mailto:arviandri064001800035@std.trisakti.ac.id), [dian.pratiwi@trisakti.ac.id](mailto:dian.pratiwi@trisakti.ac.id), [syandra\_sari@trisakti.ac.id](mailto:syandra_sari@trisakti.ac.id)3

**Abstract**

Twitter is a social media that is widely used by the people of Indonesia and the world. Twitter is also used to share news and personal opinions, market products, to criticize a policy or regulation. Opinions posted as tweets on Twitter can also be used as a benchmark for whether the policies issued are widely supported or otherwise. To obtain these benchmarks, sentiment analysis is used to separate positive opinions from negative opinions. For data retrieval, scraping from the Twitter website is used to get it. After that, the initial process is carried out before the data is processed, namely Preprocessing to eliminate parts that are not useful in data. Then the Support Vector Machine and K-Nearest Neighbor techniques were carried out to classify positive and negative opinions in order to compare which one had more opinions and then be used as a benchmark for a policy.

**Keywords**: Sentiment Analysis, UU TPKS, Support Vector Machine, K-Nearest Neighbour, Scraping

**Abstrak**

Twitter adalah media sosial yang banyak digunakan oleh masyarakat Indonesia maupun Dunia. Twitter juga dimanfaatkan untuk berbagi kabar dan opini pribadi, memasarkan produk, sampai mengkritik suatu kebijakan atau peraturan. Opini yang diposting sebagai tweet di Twitter juga dapat digunakan sebagai tolak ukur apakah kebijakan yang dikeluarkan banyak yang mendukungnya atau sebaliknya. Untuk memperoleh tolak ukur tersebut maka digunakanlah analisis sentimen untuk memisahkan opini positif dengan opini negatif. Dari pengambilan data untuk diproses maka digunakanlah scraping dari website Twitter untuk mendapatkannya. Setelah itu dilakukan proses awal sebelum data diolah yaitu Preprocessing untuk menghilangkan bagian yang tidak berguna dalam pengolahan data. Lalu dilakukan teknik Support Vector Machine dan K-Nearest Neighbour untuk mengklasifikasikan opini postif dan negatif guna untuk membandingkan manakah yang lebih banyak dari opini tersebut lalu dijadikanlah tolak ukur terhadap suatu kebijakan.

**Kata kunci**: Analisis Sentimen, UU TPKS, Support Vector Machine, K-Nearest Neighbour, Scraping

**1. PENDAHULUAN**

Akhir - akhir ini tindak pidana kekerasan seksual banyak terjadi di masyarakat, menurut Komnas Perempuan tercatat bahwa terjadi 14.719 kasus kekerasan seksual terhadap perempuan sepanjang tahun 2020. Sebenarnya pemerintah telah merancang rancangan undang – undang tindak pidana kekerasan seksual sejak tahun 2016, namun baru disahkan pada tahun 2022 setelah desakan dari berbagai pihak, disahkannya UU TPKS menimbulkan banyak pro dan kontra serta menyebabkan masyarakat beropini di media sosial. Salah satunya yang mendukung berpendapat bahwa undang – undang ini akan membuat pelaku kejahatan seksual dihukum sesuai dengan apa yang dilakukannya, sedangkan salah satu yang menentang berpendapat bahwa undang – undang ini akan menciptakan suatu pemikiran bahwa seks bebas itu diperbolehkan. Salah satu media sosial yang sering digunakan untuk beropini oleh masyarakat indonesia adalah twitter.

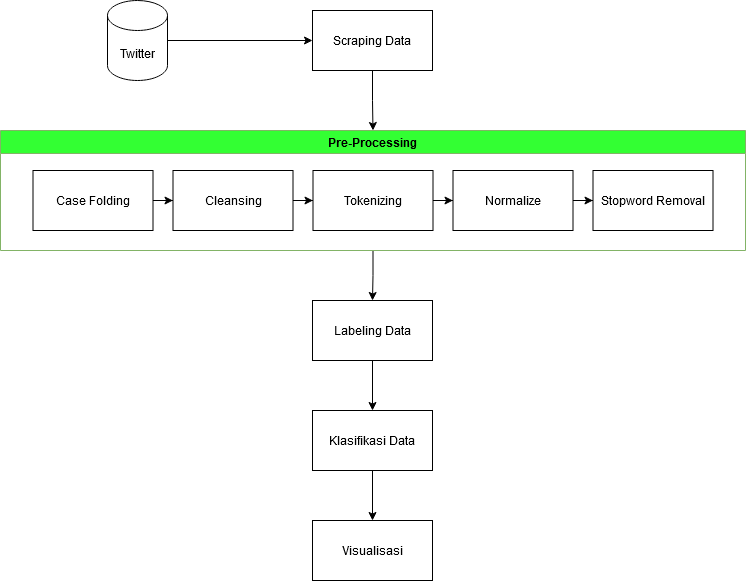
Twitter oleh masyarakat Indonesia dimanfaatkan untuk berbagai hal seperti berkomunikasi dengan orang lain secara publik atau personal, berbagi kabar dan opini pribadi, berjualan, sampai mengkritik atau memuji akan suatu hal. Dikarenakan informasi yang berada di Twitter juga dibatasi sekitar 280 karakter biasanya pengguna hanya mengirim suatu hal yang pendek [1].

Pemerintah juga memanfaatkan platform ini untuk mengetahui respon masyarakat kepada kebijakan yang baru dikeluarkan seperti pada penelitian ini. Oleh karena itu, pengguna Twitter dapat beropini tentang kebijakan yang dikeluarkan dipengaruhi oleh emosi yang dapat diklasifikasikan untuk menentukan polarisasinya, yaitu positif atau negatif tentang tweet mengenai kebijakan pemerintah pada penelitian ini.

Analisis sentimen yaitu kegiatan mengolah kata untuk menghasilkan suatu sentimen (positif atau negatif), Analisis sentimen bertujuan salah satunya yaitu untuk mendapatkan suatu opini dari kebijakan pemerintah yang baru dikeluarkan kemudian opini tersebut diklasifikasikan ke dalam sentimen positif dan negatif. Teknik yang dipakai untuk mengambil data dari Twitter sebelum di analisis yaitu menggunakan teknik Scraping yaitu mengambil data langsung dari website Twitter. Lalu teknik yang digunakan untuk memberikan sentimen (label) adalah Valence Aware Dictionary and sEntiment Reasoner (VADER) dikarenakan teknik ini memiliki akurasi yang tinggi serta bisa digunakan untuk sentiment negasi lalu teknik yang digunakan untuk mengklasifikasi data tersebut yaitu K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM) dikarenakan kedua model klasifikasi tersebut memiliki kelebihannya masing – masing pada penelitian sebelumnya tentang perbandingan akurasi keduanya yaitu tingkat akurasinya yang cukup tinggi untuk SVM (89,7 %) sedangkan KNN yaitu dapat memproses data yang besar dalam waktu singkat (1.113 data dalam waktu 0,0160 detik) [2].

**2. METODOLOGI PENELITIAN**

Pada saat menyusun tugas akhir, penulis menggunakan flowchart sebagai berikut :

****

**Gambar 1. Flowchart dari Penelitian**

Berikut langkah-langkah penyelesaian penelitian ini yaitu:

**2.1. Pengumpulan Data**

**Pengumpulan data dari salah satu media social terbesar yaitu Twitter menggunakan teknik Sraping data yang menggunakan library. Data yang dikumpulkan berupa tweet berbahasa Indonesia dengan kata kunci “uu tpks” dalam rentang waktu 12 April 2022 hingga 24 April 2022 dan tidak disertakan posting retweet.**

**2.2. Pengolahan Data**

**Setelah melakukan pengumpulan data sebelum dianalisis perlu dilakukan proses awal atau dikenal dengan istilah Preprocessing. Proses ini akan mengolah data awal yang masih tidak beraturan untuk dijadikan data teratur yang dapat diterapkan pada proses selanjutnya. Preprocessing yang dilakukan terdiri dari Case Folding, Cleansing, Tokenizing, Normalization, dan Stopword Removing.**

**2.2.1. Case Folding**

**Case Folding adalah langkah untuk melakukan perubahan huruf besar atau huruf kapital (uppercase) yang terdapat pada teks menjadi huruf kecil (lowercase).**

**2.2.2. Cleansing**

**Cleansing adalah langkah membersihkan data dari hal – hal yang tidak perlu seperti URL, hashtag, tanda baca, angka dan lain sebagainya.**

**2.2.3. Tokenizing**

**Tokenizing adalah melakukan perubahan dari suatu kata pada kalimat yang dipisahkan oleh separator (space) menjadi sebuah token.**

**2.2.4. Normalization**

**Normalization adalah suatu proses dimana kata yang tidak baku atau singkat dirubah menjadi kata baku yang benar.**

**2.2.5. Stopword Removing**

**Stopword Removing adalah proses dimana kata penghubung seperti yang, di, ke, dari yang tidak diperlukan pada proses analisis dibuang.**

**2.3 Labeling Data**

**Setelah data dibersihkan lalu dilakukan pelabelan pada data. Labeling pada data dilakukan secara otomatis menggunakan kamus yang sudah berisi bobot sentimen (lexicon) dan dihitung total dari sentimen bedasarkan jumlah bobot dari seluruh kata pada setiap data.**

**2.4 Pembobotan kata (TF-IDF)**

**Setelah di berikan label selanjutnya dilakukan pembobotan kata. Pembobotan kata dilakukan dengan menggunakan Term Frequency (TF) dan Inverse Document Frequency (IDF).**

**2.5 Mengklasifikasikan Data**

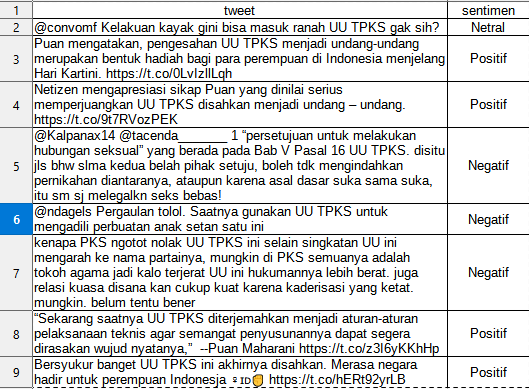
**Proses ini bertujuan untuk mengolah data menjadi opini positif dan opini negatif. Ada banyak metode untuk mengklasifikasikan data, salah satunya adalah Support Vector Machine dan K-Nearest Neighbour . Merupakan salah satu metode untuk mengklasifikasikan data dan regresi. Pada penelitian ini, penulis menggunakan metode Support Vector Machine dan K-Nearest Neighbour untuk mengklasifikasikan data.**

**2.6 Visualisasi**

**Pada proses ini akan dilakukan visualisasi terhadap data yang dihasilkan dari proses klasifikasi. Tujuan dari proses ini untuk mempermudah membaca maksud dan informasi dari hasil analisis.**

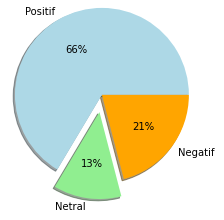
**3. HASIL DAN PEMBAHASAN**

Setelah dilakukan scrapping data menggunakan library *snscrape*, melakukan pre-processing yang terdiri dari *case folding*, *cleansing*, *tokenize*, *normalize*, dan *stopword removal* dilakukan pelabelan secara otomatis menggunakan Vader. Contoh data dari tahap ini adalah sebagai berikut :



**Gambar 2. Contoh Hasil dari Pelabelan Otomatis**

Untuk diagram pie dari data hasil pelabelan otomatis adalah sebagai berikut :



**Gambar 3. Diagram Pie Hasil Pelabelan Otomatis**

Dan wordcloud yang dihasilkan dari pelabelan otomatis adalah sebagai berikut :



**Gambar 4. Wordcloud Data Positif**

****

**Gambar 5. Wordcloud Data Negatif**

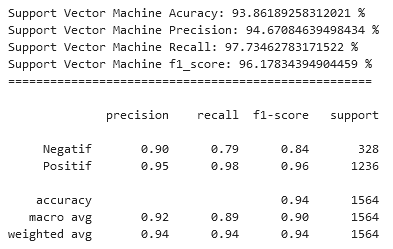
Selanjutnya dilakukan pengecekan secara manual untuk mengetahui akurasi dari Vader. Hasil dari pengecekan manual terhadap pelabelan otomatis menggunakan Vader adalah sebagai berikut :

**Tabel 1.Tabel pengecekan manual terhadap Vader**

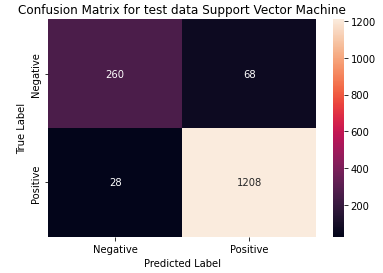
|  |  |  |
| --- | --- | --- |
| **Pengecekan Vader** | | |
| Total Data | | **1564** |
| Vader | Jumlah Positif | 1236 |
| Jumlah Negatif | 328 |
| Manual | Jumlah Positif | 1369 |
| Jumlah Negatif | 195 |
|  |  |  |
| **Vader Terhadap Manual** | | |
| Prediksi Benar | | **1403** |
| Prediksi Salah | | **161** |
| Akurasi | | **89,71%** |

Dapat dilihat dari tabel diatas bahwa akurasi yang dimiliki Vader cukup tinggi yaitu 89,71 % terhadap pelabelan manual, yaitu memiliki 1.403 prediksi yang benar dan 161 prediksi yang salah dari 1.564 data.

Setelah itu dilakukan pembobotan kata menggunakan TF-IDF dan melakukan pemodelan dan klasifikasi dengan menggunakan metode SVM dan KNN. Akurasi dan Visualisasi dari SVM tersebut adalah sebagai berikut :

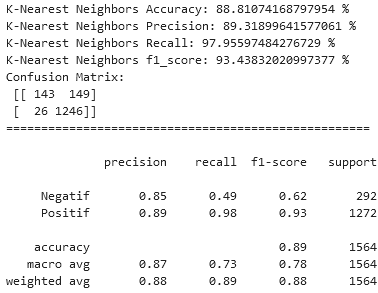
****

**Gambar 6. Akurasi dari SVM Linear**

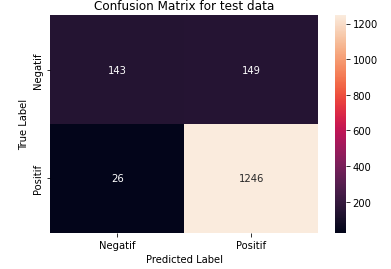
****

**Gambar 7.** **Confusion Matrix dari SVM Linear**

Sedangkan untuk akurasi dan Visualisasi dari KNN tersebut adalah sebagai berikut :



**Gambar 8. Akurasi dari KNN**

****

**Gambar 9. Confusion Matrix dari SVM Linear**

**Setelah dilakukan klasifikasi pada kedua metode tersebut lalu akurasi dari keduanya di cek menggunakan K-Fold Cross Validation untuk mengurangi bias dalam pembagian data test sebagai uji coba. Hasil yang didapatkan adalah 93,2 % untuk SVM Linear dan 87,2 % untuk KNN.**

**4. SIMPULAN**

Data yang didapatkan dalam scrapping data sebesar 15.632 data. Hasil dari sentimen yang diperoleh menggunakan Vader dalam penelitian ini yaitu positive sebesar 66 % (10.385 data), negative sebesar 21 % (3.274 data) dan netral sebesar 13 % (1.973 data). Dari hasil ini dapat disimpulkan tersebut menunjukkan kepuasan masyarakat terhadap kebijakan yang dikeluarkan tersebut.

Di dalam proses klasifikasi, ditemukan untuk klasifikasi KNN iterasi terbaik untuk penelitian ini adalah sejumlah 4 kali untuk menghasilkan akurasi yang cukup akurat. Serta kesimpulan dari klasifikasi tersebut adalah metode SVM lebih akurat dibandingkan metode KNN dengan akurasi sebesar 92 % untuk SVM dan 88 % untuk KNN.

**DAFTAR PUSTAKA**

[1] K. Makice, *Twitter API: Up and Running*. 2009.

[2] M. R. A. Nasution and M. Hayaty, “Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter,” *J. Inform.*, vol. 6, no. 2, pp. 226–235, 2019, doi: 10.31311/ji.v6i2.5129.