

Analisis Sentimen Komentar pada Media Sosial Twitter tentang PPKM Covid-19 di Indonesia dengan Metode Naïve Bayes

Aldi Bagus Sasmita¹, Bayu Rahayudi², Lailil Muflikhah³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹aldibagussasmita@gmail.com, ²ubay1@ub.ac.id, ³lailil@ub.ac.id

Abstrak

Penyebaran COVID-19 yang masih berlangsung membawa banyak perubahan, tidak terkecuali negara Indonesia. Penanganan yang tepat bagi setiap sektor untuk menghadapi pandemi ini masih terus berjalan, diantaranya adalah penetapan kebijakan untuk mencegah penyebaran virus. Upaya penetapan kebijakan yang dilakukan pemerintah untuk memutus angka penyebaran COVID-19 adalah Pemberlakuan Pembatasan Kegiatan Masyarakat yang disebut PPKM. Kebijakan yang diterapkan oleh pemerintah ini mendapat beragam tanggapan dari masyarakat melalui banyak media terutama dalam bentuk komentar. Media sosial Twitter menjadi ruang argumentasi yang efektif khususnya bagi fenomena yang sedang banyak dibicarakan termasuk di dalamnya kebijakan PPKM. Berbagai tanggapan dalam bentuk komentar perlu dilakukan analisis sentimen dengan klasifikasi tanggapan positif atau negatif yang berperan sebagai penyaring. Tujuan dilakukannya studi mengenai metode Naïve Bayes ini adalah untuk mengetahui nilai akurasi dalam klasifikasi sentimen masyarakat pada media sosial Twitter dalam menanggapi kebijakan PPKM yang dilakukan pemerintah di Indonesia. Hasil dari penelitian yang dilakukan kali ini mengutarakan bahwa algoritma *Naive Bayes Classifier* dengan menggunakan *Library Filtering NLTK* memiliki hasil akurasi yang paling tinggi dengan *Library Filtering* yang lain seperti *Library Filtering Tala* dan *Library Filtering Combined Stopword*. Nilai akurasi yang didapatkan oleh *Library Filtering NLTK* sebesar 77,2%, *Library Filtering Tala* sebesar 76,6%, dan *Library Filtering Combined Stopword* sebesar 75,2%.

Kata Kunci: Covid-19, Analisis Sentimen, PPKM, Twitter, Naïve Bayes Classifier.

Abstract

The ongoing spread of COVID-19 brings many changes, including Indonesia country. The proper handling for each sector to deal with this pandemic is still ongoing with various efforts including the establishment of policies to intercept the expansion of the virus. The government's policy-setting effort to cut the spread of COVID-19 is the Pemberlakuan Pembatasan Kegiatan Masyarakat, known as PPKM. This policy received various responses from the public through many media, especially digital media through personal social media accounts, especially in the form of comments. Twitter social media platform becomes an effective argumentation space, especially for the phenomenon that is being discussed a lot, including the PPKM policy. Various responses in the form of comments need to be analyzed by sentiment with a classification of positive or negative responses that acts as a sentence filter. The reason of this reasearch on the Naïve Bayes method is to determine the value of accuracy in the classification of public sentiment on Twitter social media in response to the PPKM policy carried out by the government in Indonesia. The consequence of the research conducted this time stated that the Naive Bayes Classifier algorithm using the NLTK Filtering Library has the highest accuracy such as Tala Filtering Library and the Combined Stopword Filtering Library. The accuracy results obtained by the NLTK Library Filtering is 77.2%, Tala Filtering Library is 76.6%, and The Combined Stopword Filtering Libray is 75.2%.

Keywords: Covid-19, Sentiment Analysis, PPKM, Twitter, Naïve Bayes Classifier.

1. PENDAHULUAN

Covid-19 ialah sebuah virus baru yang muncul dari Wuhan, China. Covid-19 ini

menyebarkan dan menyerang seluruh belahan dunia, termasuk negara Indonesia. *Covid-19* memiliki dampak yang serius bagi kehidupan. Dampak yang dirasakan ada pada faktor ekonomi serta mental manusia. Berbagai cara telah dilakukan pemerintah termasuk menerapkan program vaksin *Covid-19* dan program PPKM atau sering disebut dengan Pemberlakuan Pembatasan Kegiatan Masyarakat. Program ini menuai banyak pro dan kontra bagi masyarakat Indonesia. Beberapa masyarakat sangat terganggu terutama dengan kebijakan PPKM tersebut. Bagi mereka, PPKM membuat keterbatasan interaksi sesama manusia dengan dunia luar. Tentunya hal ini sangat membuat masyarakat tertekan dengan adanya kebijakan PPKM tersebut. Upaya masyarakat untuk menyampaikan kritik dan saran untuk pemerintah ialah melalui media sosial. Media sosial sendiri mempunyai arti yaitu sebuah media *online* yang dapat dibuka oleh semua orang (Cahyono, 2016). Salah satu dari banyaknya media sosial di Indonesia yang kini di sukai oleh banyak masyarakat adalah *Twitter* (Wiguna dan Rifai, 2021).

Twitter adalah situs *micro blogging* yang memungkinkan pengguna dapat mengirim berbagai tulisan serta membaca dan mengomentari pesan milik orang lain (Anggreini, Nasir dan Noor, 2016). Berbagai tulisan yang ditulis di media sosial *Twitter* diharapkan dapat dibaca dan ditanggapi pemerintah. Namun tulisan tersebut mempunyai banyak makna yang beragam. Terdapat tulisan yang mempunyai nilai positif ataupun negatif, namun terkadang mempunyai makna yang tidak begitu jelas atau ambigu. Hal yang dapat dilakukan untuk menyaring berbagai tulisan di *twitter* memerlukan sebuah algoritma yang digunakan untuk analisis, apakah tulisan tersebut bersifat positif atau negatif. Salah satu metode yang berguna untuk menganalisis komentar pada *twitter* ialah dengan menggunakan metode *Naïve Bayes*.

Penelitian menggunakan metode *Naïve Bayes* untuk Analisis sentimen sebelumnya pernah dilakukan oleh Fajar Ratnawati (2018) dimana penggunaan metode *Naïve Bayes Classifier* dengan pengujian pengelompokkan dataset yang digunakan menerapkan metode *5-folds cross validation* yang menghasilkan rata-rata akurasi sebesar 82,4%, precision 83,8%, Recall 82,4% dan f-measure 82,2%. Akurasi yang mempunyai nilai tinggi, menandakan bahwa sistem tersebut melakukan pencapaian

klasifikasi dengan baik. Dalam sebuah riset studi yang lain oleh Ahmad Fathan Hidayatullah (2015) menunjukkan bahwa metode SVM dan *Naïve Bayes* memiliki perbedaan akurasi. Hasil Akurasi yang didapatkan oleh metode SVM dengan proses *preprocessing stemming* memiliki nilai 92,86% dan metode *Naïve Bayes* memiliki nilai 91,96%. Lalu Hasil Akurasi yang didapatkan oleh metode SVM tanpa proses *preprocessing stemming* memiliki nilai 93,53% dan metode *Naïve Bayes* memiliki nilai 92,19% (Hidayatullah, 2015). Dalam, Penelitian lain mengenai Metode *Naïve Bayes*, juga dilakukan oleh Feng-Jen Yang (2018), Hasil penelitian yang diperoleh ialah sampel data kebenaran yang sudah diklasifikasi menggunakan *Naïve Bayes* dengan menggunakan data *Trainig* dan data *Testing*.

2. LANDASAN KEPUSTAKAAN

2.1. Analisis Sentimen

Pengertian dari analisis sentimen ialah salah satu jenis riset atau penelitian yang membahas tentang *text mining*. Terdapat perhitungan komputasional dari dua hal, yaitu opini dan perasaan yang terkandung dalam sebuah teks. Perhitungan komputasional yang dimaksud adalah analisis sentimen. Adapun harapan dari analisis sentimen ini adalah untuk mengetahui opini atau perasaan seseorang pengguna media sosial terhadap sebuah topik yang sedang hangat diperbincangkan. Positif dan negatif, ialah tanggapan atau respon yang sering dikeluarkan oleh pengguna media sosial tersebut. Ketika melakukan sebuah riset atau penelitian yang basisnya adalah analisis sentimen, kita dapat mengetahui seperti apa jawaban atau respon pengguna tersebut mengenai topik yang sedang hangat dibicarakan (Oliver, 2019).

2.2. Text Mining

Text mining sendiri mempunyai arti yang dimana proses penggalian data berupa teks, yang umumnya berupa dokumen dan mempunyai tujuan untuk mencari kata – kata yang diwakili dalam dokumen tersebut yang nantinya dapat dilakukan analisa keterkaitan yang ada dalam dokumen tersebut. Data yang berupa teks tersebut akan diproses menjadi data numerik yang nantinya dapat diproses lebih lanjut. Maka yang dimaksud *Preprocessing Data* ialah, bagian dari *Text Mining* yang dimana proses awal yang diterapkan terhadap data teks yang

nantinya berguna untuk menghasilkan data berupa numerik (Herdianto, 2018).

2.3. Twitter

Twitter adalah salah satu jenis dari sekian banyak media sosial yang banyak digunakan oleh pengguna untuk melakukan komunikasi antara satu sama lain. Pengguna twitter dapat memposting atau menuliskan tweet hanya dengan 140 karakter tweet saja. Berdasarkan portal web *techno.id*, Twitter memiliki 302 *user* atau pengguna aktif sebesar 80% yang di antaranya dari perangkat seluler. (Oliver, 2019).

2.4. Preprocessing

Guna Preprocessing sendiri ialah untuk mengolah data yang sebelumnya belum tersusun secara sistematis, lalu diolah menjadi data yang lebih sistematis. Hal ini bertujuan untuk memudahkan saat pemrosesan data. Menurut Riri Nada Devita, Heru Wahyu Herwanto, dan Aji Prasetya Wibawa (2017) ada beberapa tahapan yang dilakukan saat proses preprocessing, yaitu :

1. *Case Folding*
Case folding berguna untuk merubah data teks yang awalnya huruf acak, menjadi huruf kecil. Data harus berisi tentang A-Z, jika tidak maka tidak dapat dilakukan *case folding*.
2. *Data Cleansing*
Data Cleansing merupakan proses membersihkan dokumen dari simbol-simbol seperti link username, url dengan karakter “!”, angka dan tanda baca untuk diambil data yang akan dibutuhkan.
3. *Tokenizing*
Tahapan tokenizing merupakan proses pemotongan teks pada suatu kalimat maupun paragraf yang akan menjadi potongan yang disebut token yang nantinya akan di analisis.
4. Normalisasi
Tahapan normalisasi digunakan untuk menggaanti kata yang semula tidak baku menjadi kata baku sesuai dengan kamus yang digunakan.
5. *Filtering*
Tahapan dimana menarik kata penting dari proses Normalisasi dan menghilangkan kata yang dianggap tidak memiliki makna.
6. *Stemming*

Proses menghilangkan imbuhan yang ada di kata tersebut menjadi ke bentuk dasarnya (Devita, Herwanto dan Wibawa, 2018).

2.5. Raw Term Weighting

Raw term weighting diterapkan untuk menentukan bobot yang ada pada setiap kata yang akan di proses. Untuk menentukan raw term weighting diperlukan tahapan sebagai berikut (Nurjanah, Perdana dan Fauzi, 2017)

2.5.1 Term Frequency (TF)

Term Frequency (TF) digunakan untuk mengetahui seberapa banyak frekuensi banyaknya kemunculan term pada keseluruhan dokumen teks. Dalam pembobotan kata, Term Frequency ($tf_{t,d}$) akan menghitung berdasarkan kemunculan term t pada dokumen d . Logaritma (\log) pada metode ini berfungsi mengurangi perbedaan nilai tf yang terlalu besar. Metode ini dapat dirumuskan menggunakan persamaan sebagai berikut :

$$w_{t,d} = \begin{cases} 1 + \log_{10} tf_{t,d} & \text{if } tf_{t,d} > 0 \\ 0 & \text{if } tf_{t,d} = 0 \end{cases} \quad (1)$$

Keterangan :

$tf_{t,d}$ = jumlah kemunculan term t yang ada pada dokumen d .

2.6. Naïve Bayes Classifier

Naïve Bayes Classifier ialah sebuah metode untuk proses klasifikasi yang dapat bekerja dengan cepat dan mempunyai akurasi yang tinggi terhadap jumlah data yang banyak. Dimana asumsinya bahwa nilai *attribut* pada kelas tertentu tidak terkait pada nilai *attribut* yang lainnya (Pujadayanti, Fauzi dan Sari, 2018). Persamaan untuk menghitung metode *Naïve Bayes Classifier* menurut (Manning, et al., 2009) dapat diketahui pada persamaan sebagai berikut:

$$P(C_j|W_i) = P(C_j) \prod_{1 \leq i \leq nd} P(w_i|C_j) \quad (2)$$

Keterangan :

$P(C_j|W_i)$ = perhitungan peluang adanya kemunculan kelas C_j dimana $j = 1, 2, 3 \dots n$

$P(w_i|C_j)$ = perhitungan peluang yang ada di *term* pada kelas *j*.

$P(C_j)$ = perhitungan peluang yang ada di dokumen kelas *j* pada data latih.

2.7. Laplacian Smoothing

Metode ini dilakukan dengan cara menambahkan nilai positif di setiap probabilitasnya. Tujuan untuk menambahkan nilai positif pada probabilitas untuk menghindari hasil akhir yang bernilai null dan mempercepat proses data training pada metode Naïve Bayes (Randy, Hasniati dan Alwiah Musdar, 2018). *Laplacian smoothing* sendiri mempunyai rumus, sebagai berikut:

$$pi(y | x) = \frac{p(y|x) p(y)+k}{p(x)+k(x)} \quad (3)$$

Keterangan :

X = data yang belum teridentifikasi

y = nilai spesifik dari class

$pi(y|x)$ = probabilitas hipotesis berdasarkan kondisi

$p(y)$ = probabilitas prior

$p(x|y)$ = probabilitas kondisi berdasarkan hipotesis

$p(x)$ = probabilitas y

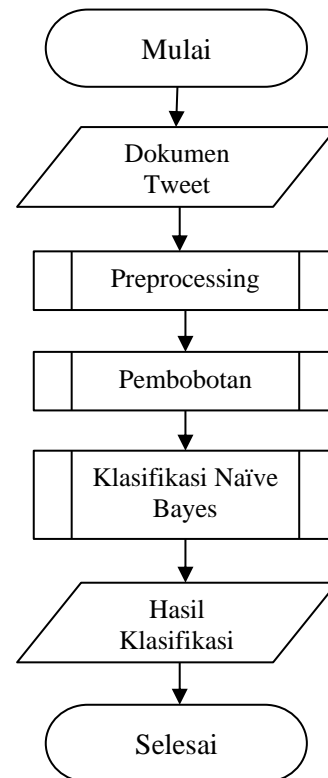
x = factor lapace smoothing

k(x) = jumlah kelas dari atribut x.

3. METODOLOGI

3.1. Diagram Alir Sistem

Berikut ialah Diagram Alir Sistem pada penelitian ini.



Gambar 1 Diagram Alir Sistem

Sistem diawali dengan memasukkan data dari dokumen *tweet*, kemudian data dilakukan *preprocessing*. Setelah dilakukan *preprocessing* kemudian dilakukan proses pembobotan dengan pembobotan *Raw Term Frequency*. Lalu selanjutnya dilakukan proses klasifikasi dengan menggunakan metode *Naïve Bayes Classifier*.

3.2. Perancangan Sistem

Pada tahap pertama, data akan diberikan label berupa positif dan negatif. Lalu tahapan selanjutnya ialah *Preprocessing* yang meliputi *Data Cleansing*, *Case Folding*, *Tokenization*, Proses Normalisasi, *Filtering* dengan 3 kamus yang berbeda, dan *Stemming*.

Data yang telah dilakukan *preprocessing*, selanjutnya melewati proses pembobotan kata, pada tahap ini dilakukan pembobotan kata dengan *Raw Term Frequency* yang berguna untuk mencari berapa banyaknya frekuensi adanya term pada keseluruhan dokumen teks. Pembobotan kata dengan menggunakan *Raw Term Frequency* kemudian akan diklasifikasikan menggunakan metode *Naïve Bayes* sehingga didapatkan hasil sentimen positif ataupun negatif.

Tahap-tahap klasifikasi menggunakan metode *Naïve Bayes* ialah:

1. Melakukan penghitungan nilai pada *Prior* yang ada pada masing-masing kelas positif dan negatif
2. Melakukan pembobotan kata menggunakan *Raw Term Frequency*
3. Menghitung nilai *Likelihood* pada setiap kelas yaitu kelas positif dan negative menggunakan rumus *Naïve Bayes* dan *Laplace Smoothing*
4. Menghitung nilai *Posterior* dari data uji untuk menentukan kelas klasifikasi
5. Menentukan kelas komentar dengan berdasarkan nilai yang paling tinggi dari *Posterior*

4. Pengujian dan Analisis

Percobaan pengujian yang dilakukan pada penelitian kali ini ialah mengukur nilai akurasi terhadap data tweets. Pada percobaan yang dilakukan kali ini, data yang digunakan adalah sebanyak 500 data yang dimana struktur dari data Positif sebanyak 50% dan Negatif 50%, dan mengambil 80% jumlah data yang dipergunakan sebagai data Latih dan 20 % jumlah data yang akan dijadikan data uji. Lalu dilanjutkan dengan pengujian menggunakan K-Fold 5.

Hasil pengujian Parameter dapat diketahui dalam tabel di bawah ini :

Tabel 1 Pengujian Parameter

No	Jumlah Data	Akurasi
1	300	75%
2	350	84.2%
3	400	78,7%
4	450	74,4%
5	500	75%

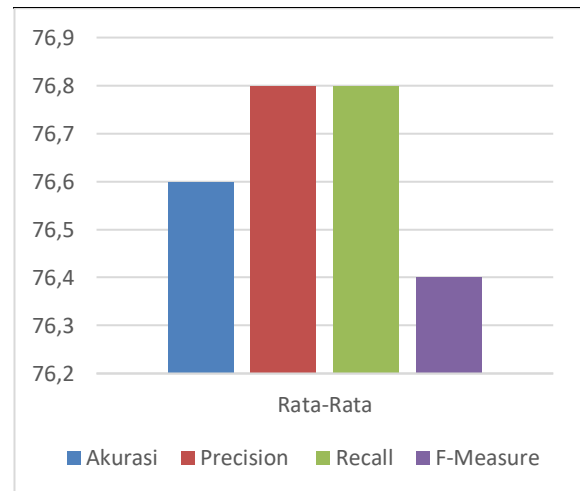
Dari hasil Parameter Pengujian Jumlah Data diatas , didapatkan hasil dengan 2 Akurasi sama yaitu sebesar 75%. Dari jumlah data 300 dan 500. Maka dari itu, data yang digunakan untuk K-Folds 5 yaitu dengan jumlah data sebesar 500 dikarenakan sebaran data positif dan negatif yang lebih beragam.

Dari hasil cross validation 5-fold dengan menggunakan 3 *Library Filtering* yang berbeda, maka didapatkan hasil rata-rata berikut :

Tabel 2 Hasil Akurasi 5 fold menggunakan *Library Filtering* Kamus Tala

Pengujian	Akurasi	Precision	Recall	f-measure
5-Fold				

Rata-rata 76,6% 76,8% 6,8% 76,4%

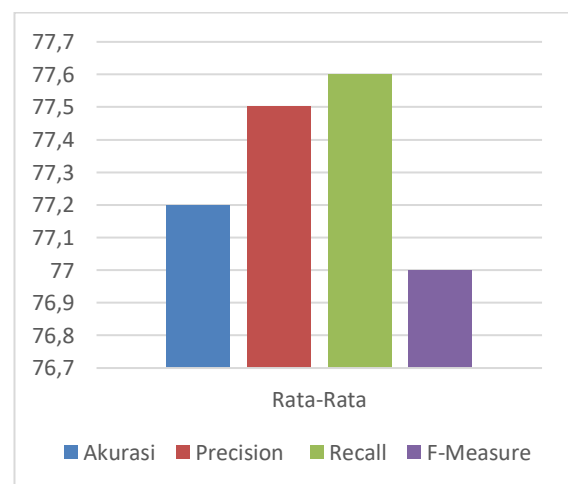


Gambar 2 Hasil Akurasi 5 fold menggunakan *Library Filtering* Kamus Tala

Tabel 3 Hasil Akurasi 5 fold menggunakan *Library Filtering* Kamus NLTK

Pengujian Akurasi Precision Recall f-measure 5-Fold

Rata-rata 77,2% 77,5% 77,6% 77%

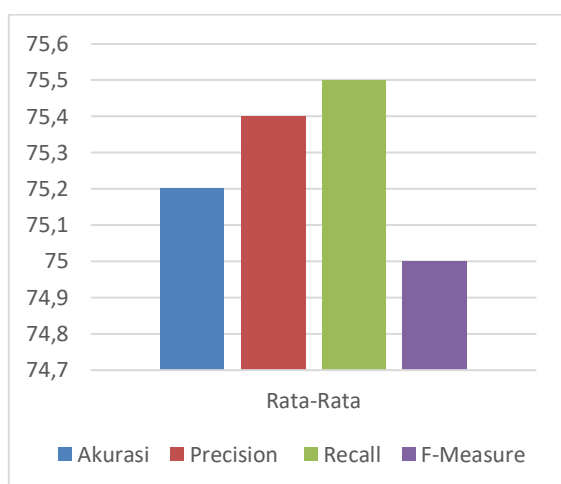


Gambar 3 Hasil Akurasi 5 fold menggunakan *Library Filtering* Kamus NLTK

Tabel 3 Hasil Akurasi 5 fold menggunakan *Library Filtering* Kamus Combined Stopword

Pengujian Akurasi Precision Recall f-measure 5-Fold

Rata-rata 75,2% 75,4% 75,5% 75%



Gambar 3 Hasil Akurasi 5 fold menggunakan *Library Filtering* Kamus *Combined Stopword*

Dari hasil keseluruhan Akurasi 5 fold pada setiap *Library* diatas dapat disimpulkan bahwa nilai akurasi setiap *Library* cenderung memiliki nilai yang beragam namun tidak berbeda jauh. Berikut keseluruhan hasil pengujian dan analisis menggunakan 5 – fold pada setiap *Library Filtering* yang terlihat pada tabel 4 berikut :

Tabel 4 Hasil Akurasi *Cross Validation* 5-fold pada setiap *Library Filtering*

<i>Library</i>	Akurasi	Precision	Recall	<i>f-measure</i>
Kamus				
Kamus	76,6%	76,8%	5,8%	76,4%
Tala				
Kamus	77,2%	77,5%	7,6%	77%
NLTK				
Kamus	75,2%	75,4%	5,5%	75%
Combined				
Stopword				

Dari hasil rata-rata keseluruhan pengujian menggunakan *cross validation* dengan 5 fold dan dengan Kamus *Filtering* yang berbeda-beda, didapatkan hasil yang cukup bervariasi namun tidak berbeda jauh antara salah satu dengan yang

lainnya. Dikarenakan, kemunculan kata yang ada pada masing-masing *library* kamus dengan jumlah isi kata yang berbeda-beda.

5. KESIMPULAN

Setelah menguji dan menganalisis hasil pada penelitian kali ini, konklusi yang dapat diambil ialah sebagai berikut:

1. Bahwa menggunakan metode *Naive Bayes* dapat menyelesaikan permasalahan analisis sentimen komentar yang ada di media sosial *Twitter* mengenai PPKM *Covid-19* dengan melalui beberapa langkah-langkah yang dimulai dari *preprocessing* , *raw term frequency*, Klasifikasi metode *Naive Bayes*, pengujian parameter, pengujian *K-Fold Cross Validation*, serta pengujian terhadap *Library* Kamus *Filtering* yang digunakan dengan nilai akurasi yang cukup beragam satu dengan yang lain, tetapi tidak cukup jauh berbeda pada setiap hasil yang dikeluarkan.
2. Metode *Naive bayes* ini terbukti dapat digunakan cukup baik untuk mengklasifikasikan data tweet di media sosial *Twitter* dari orang-orang yang membicarakan kebijakan PPKM *Covid-19*. Hal ini ditunjukkan dari hasil nilai akurasi yang didapat dari hasil percobaan yang dilakukan dengan menggunakan 5 fold. Akurasi yang didapatkan adalah 76,6% *precision* 76,8% , *recall* 76,8% dan *f-measure* 76,4% dengan menggunakan *Library* Kamus Tala, Akurasi sebesar 77,2%, *precision* 77,5%, *recall* 77,6% dan *f-measure* 77% dengan menggunakan *Library* Kamus NLTK, serta Akurasi sebesar 75,2%, *precision* 75,4%, *recall* 75,5% dan *f-measure* 75% dengan menggunakan *Library* Kamus *Combined Stopwords*

6. Saran

Untuk pengembangan riset selanjutnya, berikut beberapa saran dari penulis:

1. Perlunya penambahan jumlah dataset yang digunakan.
2. Menggunakan atau menambahkan metode lainnya.
3. Menggunakan media sosial selain twitter
4. Kategori kelas dapat ditambah lebih dari 2 kelas

DAFTAR PUSTAKA

Anggreini, N. M., Nasir, B. M., & Noor, I. L.

- (2016). Pemanfaatan Media Sosial Twitter di Kalangan Pelajar SMK Negeri 5 Samarinda. *Sosiologi*, 4(2), 239-51.
- Cahyono, A. S. (2016). Pengaruh media sosial terhadap perubahan sosial masyarakat di Indonesia. *Jurnal ilmu sosial & ilmu politik diterbitkan oleh Fakultas Ilmu Sosial & Politik, Universitas Tulungagung*, 9(1), 140–57.
- Devita, R. N., Herwanti, H. W., & Wibawa, A. P. (2018). Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa Indonesia. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 427.
- Hidayatullah, A. F. (2015). The influence of stemming on Indonesian tweet sentiment analysis.” International Conference on Electrical Engineering, Computer Science and Informatics (EECSI) 2(August). 127-32.
- Manning, c. d., Raghavan, P., & Schutze, H. (2009). an Introduction To Informational Retrieval. *Information Retrieval*.
- Herdianto, A. (2018) ‘PENCARIAN AYAT – AYAT AL-QUR’AN BERDASARKAN KONTEN MENGGUNAKAN TEXT MINING BERBASIS APLIKASI DESKTOP’, *Journal of Materials Processing Technology*, 1(1), pp. 1–8.
- Nurjanah, W. E., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIK) Universitas Brawijaya*, 1075-57.
- Oliver, J. (2019). Bab II Tinjauan Pustaka Aplikasi. *Hilos Tensados*, 1-476.
- Pujadayanti, I., Fauzi, M. A., & Sari, Y. A. (2018). Prediksi Rating Otomatis pada Ulasan Produk Kecantikan dengan Metode Naïve Bayes dan N-gram. 2(11), 4421–27.
- Randy, Hasniati and Alwiah Musdar, I. (2018) ‘Aplikasi prediksi kerusakan smartphone menggunakan metode naive bayes dan laplace smoothing’, *Jtriste*, 5(2), pp. 8–16.
- Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *INOVTEK Polbeng - Seri Informatika*, 3(1), 50.
- Wiguna, R. A. raffaidy and Rifai, A. I. (2021) ‘Analisis Text Clustering Masyarakat Di Twitter Mengenai Omnibus Law Menggunakan Orange Data Mining’, *Journal of Information Systems and Informatics*, 3(1), pp. 1–12.
- Yang, F. J. (2018). An implementation of naive bayes classifier. *Proceedings - 2018 International Conference on Computational Science and Computational Intelligence, CSCI*, (pp. 301-6).