Jurnal Sistem Informasi, Teknologi Informasi, dan Edukasi Sistem Informasi (JUST-SI)
Vol. 2, No. 1, Agustus 2021, hlm. 55-63

ANALISIS SENTIMEN KEBIJAKAN KAMPUS MERDEKA MENGGUNAKAN *NAIVE BAYES* DAN PEMBOBOTAN TF-IDF BERDASARKAN KOMENTAR PADA YOUTUBE

Dhaifa Farah Zhafira[*1], Bayu Rahayudi[2], Indriati[3]

[1,2,3]Universitas Brawijaya
Email: [1] dhaifafarah@student.ub.ac.id, [2] ubay1@ub.ac.id, [3] indriati.tif@ub.ac.id
[*]Penulis Korespondensi

(Naskah masuk: 13 Januari 2021, diterima untuk diterbitkan: 18 Agustus 2021)

Abstrak

Kebijakan Kampus Merdeka merupakan salah satu kebijakan baru yang digagas oleh Menteri Pendidikan dan Kebudayaan Republik Indonesia (Mendikbud RI). Kebijakan tersebut tengah ramai disorot publik khususnya pada platform Youtube berkaitan dengan video unggahan Mendikbud di kanalnya. Pada Youtube, opini masyarakat dapat membanjiri kolom komentar dalam sekejap karena kemunculannya sebagai *platform* pertama yang menawarkan fasilitas konten audio visual. Penelitian ini mencoba menganalisis opini masyarakat yang tertampung dalam kolom komentar Youtube ke dalam klasifikasi sentimen positif dan negatif. Klasifikasi diimplementasikan pada Google *Colaboratory* yang berbasis bahasa Python dan Jupyter Notebook dengan algoritme *Naive Bayes Classifier* serta pembobotan kata *Term Frequency Inverse Document Frequency* (TF-IDF). 5 proses utama dalam penelitian ini yang meliputi pelabelan manual, *text preprocessing*, pembobotan TF-IDF, validasi data menggunakan *k-fold cross validation*, dan klasifikasi. Hasil akurasi terbaik sebesar 97% yang didapat dengan menggunakan 900 data latih, 100 data uji, menerapkan pembobotan TF-IDF, dan 10-*fold cross validation*. Rata-rata akurasi yang didapat dari 10 iterasi pada *k-fold cross validation* yaitu sebesar 91.8% dengan nilai *precision, recall, f-measure* sebesar 90.35%, 93.6%, 91.95%. Berdasarkan hasil tersebut, *Naive Bayes Classifier* cukup baik sebagai alternatif untuk analisis sentimen.

Kata kunci: *analisis sentimen, kampus merdeka, naive bayes, TF-IDF, k-fold cross validation, youtube*

SENTIMENT ANLAYSIS OF KAMPUS MERDEKA POLICY USING NAIVE BAYES AND TF-IDF TERM WEIGHTING BASED ON YOUTUBE COMMENTS

Abstract

The Policy of Kampus Merdeka is one of the new policies initiated by the Minister of Education and Culture Republic of Indonesia (Mendikbud RI). This policy has been under public spotlight, especially on the Youtube platform and connected with the uploaded video by the Minister of Education and Culture channel. On the Youtube platform, public opinion can flood the comments column in an instant because currently Youtube is one of the biggest platform because of its appearance as the first platform offers audio-visual content. This reserach tries to analyze public opinion that is accommodated in the Youtube comments column into the classification of positive and negative sentiments. The classification process

is implemented in Google Collaboratory based on Python and Jupyter Notebook using Naive Bayes Classifier algorithm and term weighting Term Frequency Inverse Document Frequency (TF-IDF). 5 main processes in this reserach included manual labeling, text preprocessing, TF-IDF weighting, data validation using k-fold cross validation, and classification. The best accuracy results is 97% were obtained using 900 training data, 100 test data, using TF-IDF weighting, and 10-fold cross validation. The average accuracy obtained from 10 iterations on k-fold cross validation is 91.8% with precision, recall, f-measure is 90.35%, 93.6%, 91.95%. From that results, Naive Bayes Classifier is good enough for sentiment analysis alternative.

Keywords: *sentiment analysis, kampus merdeka, naive bayes, TF-IDF, k-fold cross validation, youtube*

1. PENDAHULUAN

Perkembangan teknologi membawa banyak keuntungan, salah satunya yaitu kemudahan dalam mengakses informasi melalui internet. Hal tersebut diiringi dengan semakin banyak tumbuhnya *platform* digital seperti Facebook, Instagram, Twitter, Youtube, dan banyak lagi. Hadirnya Youtube menjadi inovasi baru karena informasi yang disajikan dalam bentuk audio visual. Pengaruh cepatnya laju penyebaran informasi dan kepopuleran Youtube saat ini dimanfaatkan banyak pihak sebagai wadah untuk *branding* dan publikasi, termasuk instansi pemerintah. Kementerian Pendidikan dan Kebudayaan (Kemendikbud) merupakan salah satu instansi pemerintah yang memanfaatkan Youtube untuk penyebaran informasi, seperti halnya unggahan Kemendikbud pada kanal resmi mengenai peluncuran Kebijakan Kampus Merdeka.

Kebijakan Kampus Merdeka diluncurkan dengan tujuan menunjang persiapan kompetensi mahasiswa untuk menjawab kebutuhan zaman seiring perubahan sosial, budaya, dunia kerja, dan teknologi yang amat signifikan. Kebijakan ini cukup menjadi sorotan sejak awal kemunculannya, menuai pro dan kontra. Dengan adanya *platform* digital yang mampu mewadahi opini masyarakat, pro maupun kontra bukan hanya bermunculan dari layar kaca melalui liputan berita tetapi juga banyak bermunculan di media sosial dan *platform* digital seperti Youtube. Dalam waktu sekejap ratusan bahkan ribuan opini masyarakat memenuhi kolom komentar pada video unggahan Kemendikbud mengenai kebijakan tersebut. Hal ini tentu memerlukan waktu yang lama untuk mengelompokkan sentimen masyarakat secara manual. Oleh karena itu, peneliti menerapkan *machine learning* menggunakan algoritme *Naive Bayes Classifier* untuk klasifikasi sentimen terkait Kebijakan Kampus Merdeka.

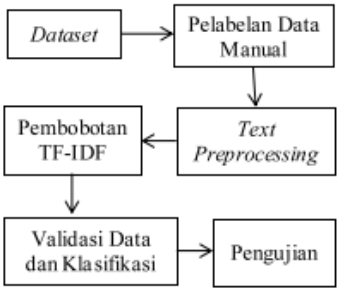
Algoritme *Naive Bayes Classifier* banyak digunakan pada penelitian terdahulu untuk menganalisis klasifikasi sentimen. Menurut Hamzah (2012), *Naive Bayes Classifier* memiliki banyak keunggulan salah satunya yaitu cepat dalam perhitungan, algoritme yang sederhana, dan dapat menghasilkan akurasi tinggi. Penelitian sebelumnya yang menggunakan *Naive Bayes Classifier* merupakan penelitian mengenai sentimen masyarakat terhadap *transgender* berdasarkan komentar di Instagram dan menghasilkan akurasi sebesar 93,33% (Putra, 2019). Pada penelitian ini menerapkan pembobotan *Term Frequency Inverse Document Frequency* (TF-IDF) dan seleksi fitur *chi-square*, tetapi akurasi tertinggi didapatkan tanpa adanya penerapan seleksi fitur *chi-square*. Penelitian lainnya yaitu mengenai opini film di Twitter yang juga menggunakan *Naive Bayes Classifier*, perbaikan kata tidak baku, dan pembobotan kata (TF saja). Pada penelitian ini didapat akurasi terbaik sebesar 91,67% (Antinasari, 2017). Berdasarkan pemaparan di atas menunjukkan bahwa *Naive Bayes Classifier* cukup baik untuk analisis sentimen mengenai Kebijakan Kampus Merdeka pada penelitian ini.

Penelitian ini menerapkan beberapa hal meliputi tahap pelabelan data secara manual, *text preprocessing* yang mencakup perbaikan kata tidak baku dengan kamus yang disusun oleh penulis, pembobotan kata *Term Frequency Inverse Document Frequency* (TF-IDF), kemudian

dilakukan validasi data menggunakan *k-fold cross validation* untuk mendapatkan bukan hanya hasil yang baik tetapi juga valid. Selain itu juga dilakukan uji variasi jumlah data latih untuk mendapatkan komposisi data latih dan data uji yang baik. Proses pengujian menggunakan metode *confusion matrix* dengan mengacu pada skor *accuracy, precision, recall, f-measure* yang dihasilkan. Melalui penelitian ini diharapkan analisis sentimen yang dihasilkan dapat membantu para pemangku kebijakan maupun pihak yang berkaitan untuk mengelompokkan sentimen masyarakat dengan metode yang lebih efektif dan akurat tanpa harus melakukannya secara manual yaitu dengan menggunakan *Naive Bayes Classifier.*

2. METODOLOGI PENELITIAN

Algoritme yang digunakan dalam penelitian yaitu algoritme *Naive Bayes Classifier* dengan adanya perbaikan kata tidak baku pada tahap *text preprocessing*, ditambah pembobotan kata TF-IDF dan validasi model menggunakan *k-fold cross validation* seperti direpresentasikan pada Gambar 1. Klasifikasi sentimen diimplementasikan di Google *Colaboratory* dengan bahasa Python dan Jupyter Notebook sifatnya *online-based service*.



Gambar 1. Metodologi Penelitian

2.1. *Dataset*

Data yang digunakan dalam penelitian merupakan data yang tersedia secara publik pada *platform* Youtube. Teknik pengumpulan data yang digunakan merupakan teknik *scraping*. Teknik *scraping* merupakan teknik untuk mendapatkan informasi dari sebuah laman secara otomatis tanpa harus menyalin secara manual (Ayani dkk., 2019). Proses *scraping* dapat dilakukan melalui Google *Spreadsheet* yang terintegrasi dengan Google *Apps Script* untuk menuliskan baris perintah *scraping* pada halaman kerjanya. Proses *scraping* data yang bersumber dari Youtube dapat memanfaatkan fasilitas Youtube Data API yang sudah tersedia pada menu *resources*. Dataset keseluruhan yaitu sebanyak 1000 data komentar dengan komposisi yang seimbang antara sentimen positif dan negatif.

2.2. Pelabelan Data

Proses pelabelan data secara manual dilakukan oleh 3 aktor dengan latar belakang bidang yang berbeda dan masih ada keterkaitan dengan Kebijakan Kampus Merdeka. Pelabelan oleh 3 aktor dimaksudkan untuk menghindari unsur subyektif pada label sentimen yang diberikan. 3 aktor tersebut merupakan pengguna Youtube berlatarbelakang bidang Ilmu Bahasa, Ilmu Psikologi, dan Ilmu Komputer (penulis). Teknik pelabelan data ditunjukkan seperti pada Tabel 1 yaitu jumlah dominan sentimen dari 3 aktor akan menjadi hasil akhir label sentimen.

Tabel 1. Pelabelan Data Manual

Data	1	2	3	Hasil
Kita liat aja ntar pelaksanaannya	pos	neg	pos	pos

2.3. *Text Mining*

Feldman dan Sanger (2007) menyatakan bahwa *text mining* bisa diartikan seperti sebuah proses mengekstrak informasi berbentuk teks, sedangkan informasi tersebut adalah kecenderungan dalam bentuk pola statistik.

2.3.1. *Text Preprocessing*

Putra (2018) menyatakan bahwa *text preprocessing* digunakan untuk transformasi tata data menjadi lebih terstruktur. Dalam penelitian ini *text preprocessing* terbagi atas c*ase folding, tokenization,* perbaikan kata tidak baku, *stopwords removal,* dan *stemming*.

Case folding merupakan proses perubahan karakter pada data alam bentuk transformasi semua huruf pada data menjadi huruf kecil, sedangkan selain huruf akan dihilangkan (Indraloka & Santosa, 2017). *Tokenization* merupakan proses pemisahan teks menjadi bagian-bagian tertentu yang biasa disebut dengan *token* (Indraloka & Santosa, 2017). Pada penelitian ini *token* yang menjadi keluaran dari proses *tokenization* berupa kata. Perbaikan kata tidak baku merupakan proses dilakukannya perubahan pada kata-kata dalam teks yang belum baku dan ditransformasi menjadi kata yang baku. Kamus perbaikan kata pada penelitian ini disusun oleh penulis. *Stopword removal* merupakan tahapan penyaringan kata-kata penting dari data yang didapat. Kata yang ada di dalam *list stopwords* akan dihapus karena dianggap tidak memengaruhi hasil analisis sentimen. *Stemming* merupakan proses pemetaan berbagai variasi dari morfologi kata yang dikembalikan ke bentuk dasarnya (Indraloka & Santosa, 2017).

2.4. Pembobotan *Term Frequency Inverse Document Frequency* (TF-IDF)

Metode *Term Frequency Inverse Document Frequency* (TF-IDF) ialah teknik penentuan seberapa *term* mewakili konten dalam dokumen dengan memberi bobot ke masing-masing kata yang terkandung di dalamnya (Karmayasa & Mahendra, 2010). Nilai TF-IDF didapat dari perkalian TF dan IDF. Nilai TF didapat dari perhitungan Persamaan 1. Nilai IDF didapat dari perhitungan Persamaan 2. Dengan $tf_{t,d}$ merupakan nilai *term frequency* t di dokumen d, $N_{t,d}$ merupakan jumlah munculnya *term* t di dokumen d, N_d merupakan total *term* yang terdapat pada dokumen d, idf_t merupakan nilai IDF dari *term* t, n merupakan jumlah koleksi dokumen, dan n_k merupakan banyaknya dokumen yang memuat *term* t.

$$tf_{t,d} = \frac{N_{t,d}}{N_d} \quad (1)$$

$$idf_t = \log \frac{n}{n_k} \quad (2)$$

2.5. *Naive Bayes Classifier*

Naive Bayes Classifier merupakan metode dalam penambangan data yang populer dikarenakan kemudahan penggunaannya (Hall, 2006), kecepatan waktu proses, kemudahan implementasi dengan strukturnya yang sederhana, serta efektivitasnya tinggi (Taheri & Mammadov, 2013). Konsep *Naive Bayes* merupakan bentuk prediksi peluang di masa mendatang berdasarkan pengalaman yang terjadi di masa sebelumnya. Perhitungan untuk klasifikasi menggunakan *Naive Bayes* dapat menggunakan Persamaan 3. $P(j|t)$ merupakan probabilitas kemunculan *term* t pada kategori j, $P(t|j)$ merupakan probabilitas *term* t

berkategori j, $P(j)$ merupakan probabilitas dokumen berkategori j, dan $P(t)$ merupakan probabilitas munculnya *term* t.

$$P(j|t) = \frac{P(j)P(t|j)}{P(t)}$$

(3)

2.6. *K-Fold Cross Validation*

Menurut Tempola dkk. (2018), *cross validation* merupakan metode yang dimanfaatkan untuk menelisik keberhasilan suatu model dengan menerapkan perulangan dengan komposisi setiap data pernah menjadi data latih maupun data uji sehingga algoritme yang digunakan teruji validitasnya. Sedangkan *k-fold cross validation* sendiri dapat didefinisikan sebagai teknik validasi silang yang menerapkan pemecahan data ke dalam k sub-set data dengan pembagian jumlah yang seimbang. Pada prosesnya dilakukan pelatihan dan pengujian yang iteratif sebanyak k kali. Data dipecah menjadi k bagian.

2.7. *Confusion Matrix*

Confusion matrix merupakan sebuah alat yang digunakan untuk mengukur seberapa baik sebuah *classifier* yang digunakan dalam mengenali tuple dari kelas yang berbeda (Han & Kamber, 2011). Pengukuran yang diterapkan *confusion matrix* yaitu dengan menghitung *accuracy, precision, recall*, *f-measure* yang mengacu pada nilai *True Positive* (TP), T*rue Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) yang merupakan keluaran dari *confusion matrix*. Parameter kinerja *classifier* mengacu pada hasil *accuracy* apabila selisih tipis antara nilai FP dan FN.

3. HASIL DAN PEMBAHASAN

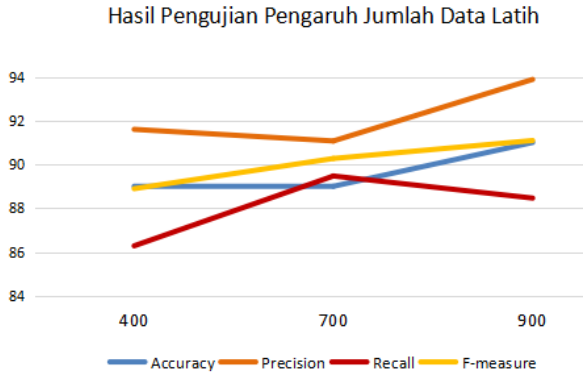
Penelitian ini terbagi menjadi 3 proses pengujian yang terdiri atas pengujian pengaruh jumlah data latih yang digunakan, pengujian pengaruh pembobotan TF-IDF, dan validasi data menggunakan *k-fold cross validation*. Pengujian dilakukan menggunakan metode c*onfusion matrix* yang keluarannya akan dimasukkan ke dalam perhitungan *accuracy, precision, recall,* dan *f-measure* sebagai *score* pengujian.

3.1. Pengujian Pengaruh Jumlah Data Latih

Pengujian pengaruh jumlah data latih dilakukan dengan memberikan variasi jumlah data latih yang digunakan dalam klasifikasi. Jumlah data latih yang diuji yaitu 400, 700, dan 900. Sedangkan jumlah data uji dibuat konstan sebesar 100 data. Data yang digunakan dalam pengujian ini merupakan data yang telah melalui proses *text preprocessing*. Tabel 9 dan Gambar 3 menunjukkan hasil pengujian pengaruh jumlah data latih.

Tabel 2. Hasil Pengujian Pengaruh Jumlah Data Latih

Jumlah Data Latih	*Accuracy*	*Precision*	*Recall*	*F-measure*
400	89%	91.6%	86.28%	88.89%
700	89%	91.07%	89.47%	90.27%
900	91%	93.88%	88.46%	91.09%



Gambar 2. Pengaruh Jumlah Data Latih

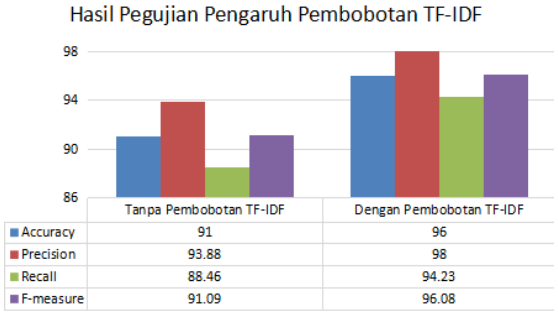
Tabel 9 dan Gambar 3 merepresentasikan bahwa jumlah data latih terbukti berpengaruh terhadap hasil klasifikasi. Penggunaan 400 data latih menghasilkan accuracy sebesar 89%. Penggunaan 700 data latih menghasilkan accuracy sebesar 89%. Penggunaan 900 data latih mendapat hasil akurasi terbaik yaitu sebesar 91%. Dari perubahan hasil pengujian dari jumlah data latih sebesar 400 hingga 900 dapat disimpulkan bahwa semakin besar penggunaan data latih cenderung meningkatkan nilai *accuracy*. Jumlah data latih terbaik dari pengujian ini akan digunakan sebagai porsi data latih pada pengujian selanjutnya.

3.2. Pengujian Pengaruh Pembobotan TF-IDF

Pengujian ini terdiri atas 2 proses yaitu melakukan klasifikasi tanpa pembobotan TF-IDF dan dengan menerapkan pembobotan TF-IDF. Data yang digunakan merupakan data dengan hasil terbaik pada pengujian sebelumnya. Tabel 3 dan Gambar 3 menunjukkan hasil pengujian pengaruh pembobotan TF-IDF terhadap hasil klasifikasi.

Tabel 3. Pengujian Pengaruh Pembobotan TF-IDF

TF-IDF	***Accuracy***	***Precision***	***Recall***	***F-measure***
Tidak	91%	93.88%	88.46%	91.09%
Ya	96%	98%	94.23%	96.08%

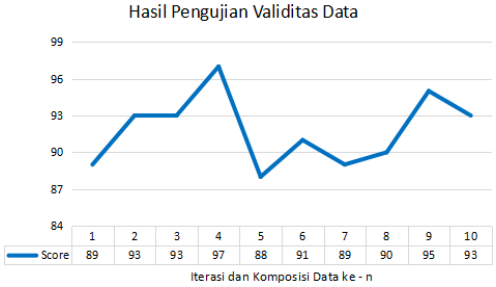


Gambar 3. Pengujian Pengaruh Pembobotan TF-IDF

Tabel 3 dan Gambar 3 menunjukkan bahwa pembobotan TF-IDF berpengaruh terhadap hasil klasifikasi. Pada proses tanpa diberi perlakuan TF-IDF menghasilkan *accuracy* 91% sedangkan proses yang menggunakan TF-IDF menghasilkan *accuracy* sebesar 96%. Dari hasil tersebut dapat disimpulkan bahwa dengan perlakuan pembobotan TF-IDF dapat meningkatkan nilai *accuracy*.

3.2. Pengujian Validitas Data

Pengujian validitas data dimaksudkan untuk mengetahui validitas data menggunakan metode *k-fold cross validation*. Nilai k yang digunakan adalah k=10. Sehingga pada pengujian ini menghasilkan 10 *score* keluaran setiap iterasinya. Data yang digunakan merupakan data dengan hasil terbaik pada pengujian sebelumnya. Tabel 4 dan Gambar 4 merepresentasikan hasil pengujian validitas data menggunakan *k-fold cross validation*.



Gambar 4. Pengujian Validitas Data

Tabel 4. Pengujian Validitas Data

Iterasi ke -	***Score***
1	89%
2	93%
3	93%
4	97%
5	88%
6	91%
7	89%
8	90%
9	95%
10	93%
Rata-rata	**91.8%**

Pengujian ini menghasilkan nilai *precision, recall, f-measure* sebesar 90.35%, 93.6%, 91.95%. Sedangkan nilai *accuracy* keseluruhan pengujian validitas data didapat dari rata-rata setiap iterasi yaitu sebesar 91.8%. *Score accuracy* tertinggi terjadi pada komposisi data di iterasi ke-4 sebesar 97%. Pada komposisi tersebut potongan data ke-4 menjadi data uji sedangkan data lainnya sebagai data latih.

4. KESIMPULAN

Proses klasifikasi sentimen menggunakan *Naive Bayes Classifier* yang diterapkan pada penelitian ini terdiri atas pelabelan data secara manual oleh 3 aktor, *text preprocessing*, pembobotan TF-IDF, kemudian proses klasifikasi itu sendiri. Proses di dalam *text preprocessing* yang diterapkan meliputi *case folding, tokenization*, perbaikan kata tidak baku, *stopwords removal,* dan *stemming*. Pada proses perbaikan kata tidak baku menggunakan kamus perbaikan kata yang dibuat oleh penulis. Setelah proses *text preprocessing* dilakukan pembobotan *term* menggunakan TF-IDF dan dilanjutkan ke proses klasifikasi menggunakan algoritme *Naive Bayes Classifier*. *Dataset* yang digunakan merupakan data dengan komposisi seimbang antara sentimen positif dan negatif dari hasil *scraping* komentar pada Youtube yang dimaksudkan untuk meminimalisir bias dalam pengujian *classifier* yang digunakan.

Hasil dari beberapa pengujian yang diterapkan pada penelitian membuktikan beberapa hal di bawah ini:

a. Semakin banyak data latih yang digunakan cenderung meningkatkan nilai *accuracy* dari hasil klasifikasi
b. Penerapan pembobotan TF-IDF berpengaruh pada peningkatan *accuracy* dari hasil klasifikasi yang mulanya 91% menjadi 96%.
c. Hasil terbaik pada penelitian ini sebesar 97% didapat dari iterasi ke-4 pada *10-fold cross validation* dengan komposisi data latih sebesar 900 dengan 100 data uji dan penerapan pembobotan TF-IDF. Rata-rata *score accuracy* yang didapat dari uji validitas data menggunakan *k-fold cross validation* yaitu 91.8% dengan nilai *precision, recall, f-measure* sejumlah 90.35%, 93.6%, 91.95%.

Dari hasil tersebut menunjukkan *Naive Bayes Classifier* dengan beberapa improvisasi dan penambahan dapat dikategorikan mendapati hasil yang sangat baik. Persebaran kata yang didapat dari hasil klasifikasi sentimen negatif didominiasi oleh kata 'kampus merdeka', 'kebijakan', 'magang' sehingga dapat disimpulkan secara umum bahwa komentar negatif mengindikasikan kurang setujunya terkait kebijakan terutama pemberlakuan magang. Sedangkan persebaran kata yang didapat dari hasil klasifikasi sentimen negatif didominiasi oleh kata 'menteri', 'memotivasi', 'bagus' yang dapat disimpulkan secara umum bahwa komentar positif mengindikasikan dukungan untuk ide menteri pencetus karena kebijakan yang diusung memotivasi bahkan dinilai 'bagus'.

DAFTAR PUSTAKA

Antinasari, P. 2017. Analisis Sentimen tentang Opini Film pada Dokumen Twitter berbahasa Indonesia menggunakan Naive Bayes Classifier denan perbaikan Kata Tidak Baku. S.Kom Thesis. Universitas Brawijaya.

Ayani, D. D., Pratiwi, H. S., Muhardi, Hafiz. 2019. Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace. Jurnal Sistem dan Teknologi Informasi, Volume VII, pp. 257-262.

Feldman, R., & Sanger, J. 2007. The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. New York: Cambridge University Press.

Hall, M. (2006). A Decision Tree-Based Attribute Weighting Filter for Naive Bayes. Knowledge-Based Systems, 20(2), 59-70.

Hamzah, A. 2012. Klasifikasi Teks dengan Naive Bayes Classifier untuk Pengelompokan Teks Berita dan Abstract Akademis. Jurnal Prosiding Seminar nasional Aplikasi & Teknologi (SNAST) Periode III.

Han, J., & M. Kamber. 2001. Data Mining: Concepts and Techniques. San Francisco: Morgan Kaufman Publisher.

lndraloka, D. S., & Santosa, B. 2017. Penerapan Text Mining untuk Melakukan Clustering Data Tweet Shopee Indonesia. Jurnal Sains dan Seni ITS Vol. 6.

Karmayasa, O., & Mahendra, I. B. 2010. Implementasi Vector Space Model dan Beberapa Notasi Metode Term Frequency Inverse Document Frequency (TF-IDF) pada Sistem Temu Kembali Informasi. Jurnal Program Studi Teknik Informatika Universitas Udayana.

Putra, E., 2019. Klasifikasi Sentimen Masyarakat terhadap Transgender berdasarkan Komentar di Instagram menggunakan Metode Naive Bayes Classifier. S.Kom Thesis. Universitas Islam Negeri Sultan Syarif Kasim Riau.

Taheri, S. & Mammadov, M. 2013. Learning the Naive Bayes Classifier with Optimization Models. International Journal of Applied Mathematics and Computer Science, 878-785.

Tempola, F., Muhammad, M., & Khairan A. 2018. Perbandingan Klasifikasi antara KNN dan Naive Bayes pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation. Jurnal Teknologi Informasi dan Ilmu Komputer, 577-584.