

Analisis Sentimen Terhadap Undang-Undang Cipta Kerja Menggunakan Metode *Naïve Bayes*

I Gusti Naufhal Daffa Adnyana¹, Fikri Adams², Anggun Windari Oktavia³, Ermatita⁴, Sarika⁵
S1 Informatika / Fakultas Ilmu Komputer

Program Studi Informatika, Universitas Pembangunan Nasional Veteran Jakarta
Jl. RS. Fatmawati Raya, Pd. Labu, Kec. Cilandak, Kota Depok, Jawa Barat 12450 igustinda@upnvj.ac.id ,
fikriadams@upnvj.ac.id , windarianggun@gmail.com , ermatitaz@yahoo.com, sarika.afrizal@upnvj.ac.id

Abstrak. *Twitter* adalah platform yang memungkinkan orang untuk mengekspresikan keinginan, pendapat, dan kritik mereka secara langsung. Disahkannya dan diundangkannya UU Cipta Kerja ini menimbulkan banyak perbedaan pendapat di masyarakat, khususnya pengguna *Twitter*. Tujuan dari penelitian ini adalah untuk memahami apa yang dipikirkan pengguna *Twitter* tentang Undang-Undang Cipta Kerja melalui hashtag #UUCIPTAKERJA. Penelitian ini membagi opini menjadi positif dan negatif, kemudian menggunakan algoritma *Naïve Bayes* untuk mengklasifikasikan sentiment pada *tweet*. Data diperoleh dengan menggunakan API yang disediakan oleh *Twitter* dari 5 Oktober hingga 30 November 2020. Keakuratan hasil pengujian yang diperoleh dengan menggunakan data uji adalah 80,53%, nilai recall 84,78%, dan nilai spesifisitas 73,79%.

Kata Kunci: *Twitter*, Undang-Undang Cipta Kerja, Klasifikasi, *Naïve Bayes*

1 Pendahuluan

Twitter adalah salah satu media sosial yang paling populer untuk bertukar informasi. Informasi yang dibagikan dapat berupa berita atau opini, dan opini baik positif maupun negatif dapat dimasukkan dalam bentuk tweet di *Twitter*. Pendapat tersebut dapat diarahkan pada suatu isu, topik atau institusi dan organisasi tertentu, dan publik dapat memberikan tweet berupa evaluasi atau tanggapan terhadap topik yang sedang dibahas.

Tentang undang-undang penciptaan lapangan kerja sebagai objek penelitian ini. Menurut Pasal 3 Undang-Undang Nomor 11 Tahun 2020 tentang Penciptaan Lapangan Kerja, disebutkan bahwa tujuan undang-undang ini adalah untuk menciptakan dan meningkatkan lapangan kerja dengan mempromosikan dan memberdayakan UMKM untuk menyerap tenaga kerja seluas-luasnya, memastikan bahwa setiap warga negara dipekerjakan dan diberikan kesempatan kerja, upah dan perlakuan yang adil, penyesuaian koperasi, UMKM, dan industri nasional, serta memperbaiki masalah pada penanaman modal[1]. Undang-undang ini banyak mendapat penolakan dari masyarakat karena dinilai merugikan para pekerja dimana beberapa isu yang menjadi masalah adalah pemotongan pesangon, pembatalan cuti melahirkan, dan terlalu cepatnya proses pembuatan undang-undang karena Presiden Joko Widodo hanya memberikan tenggat waktu 100 hari untuk menyelesaikan undang-undang tersebut sehingga terkesan terburu-buru, dan sedikitnya jumlah pihak yang dilibatkan dalam pembuatan undang-undang[2].

Penelitian mengenai analisis sentimen telah banyak dilakukan sebelumnya, Wijaya, dkk [3]. telah melakukan penelitian serupa terhadap Undang-Undang Cipta Kerja dengan mengklasifikasikan opini dari *tweet* yang dimiliki dengan menggunakan *naïve bayes classifier* yang mana mendapatkan nilai akurasi sebesar 89,9%, *precision* sebesar 90%, *recall* sebesar 89,9%, dan *f-1 score* sebesar 89,9% dimana sebanyak 52,9% masyarakat menolak Undang-Undang Cipta Kerja dan hanya 47,1% saja yang mendukung Undang-Undang Cipta Kerja. Kemudian ada Sandryan, dkk [4]. yang melakukan penelitian juga terhadap Undang-Undang Cipta Kerja dengan menggunakan algoritma *backpropagation* dan *term frequency-inverse document frequency* dengan hasil akurasi sebesar 95%, *precision* sebesar 98%, *recall* sebesar 92,4%, dan *f-measure* sebesar 95,1%.

Analisis sentimen dilakukan untuk membantu memahami opini dan tanggapan orang dalam bentuk teks. Analisis sentimen berfokus pada opini atau tanggapan yang secara langsung maupun tidak langsung memiliki sentimen positif dan negatif, di mana *text mining* digunakan saat memproses teks yang fungsinya untuk mengubah teks tidak terstruktur menjadi terstruktur terlebih dahulu. Dalam proses pengklasifikasian sentimen ke dalam kategori positif dan negatif, penulis menggunakan salah satu metode klasifikasi yaitu *Naive Bayes* (NB).

Algoritma *Naive Bayes* (NB) yang digunakan untuk klasifikasi sentimen pada penelitian ini diharapkan dapat

memberikan model klasifikasi yang baik, sehingga informasi yang diperoleh dari hasil model klasifikasi tersebut kredibel dan dapat digunakan untuk analisis sentiment terhadap reaksi masyarakat Indonesia dalam hal ini para warganet pengguna *twitter* tentang Undang-Undang Cipta Kerja

2 Metodologi Penelitian

2.1 Analisis Sentimen

Analisis sentimen atau opinion mining adalah studi yang menganalisis opini dan sentimen publik terhadap entitas seperti produk dan layanan yang disediakan dalam bentuk teks[5]. Analisis sentimen berfokus pada opini yang mengungkapkan atau menyarankan sentimen positif dan negatif[6].

2.2 Text Mining

Penambangan teks (*text mining*) dapat dianggap sebagai mengakses lebih banyak informasi secara mendalam untuk membantu pengguna menganalisis dan mencerna informasi yang dapat membantu dalam pengambilan keputusan. *Text mining* digunakan saat memproses teks yang fungsinya untuk mengubah teks tidak terstruktur menjadi terstruktur terlebih dahulu. Tujuan utama dari *text mining* adalah untuk menganalisis informasi untuk mendapatkan pola dari informasi tersebut[7].

2.3 Text Analysis

Analisis teks adalah metode dan proses yang diikuti untuk memperoleh informasi dan wawasan berkualitas tinggi dan dapat ditindaklanjuti dari data teks. Analisis teks terdiri dari serangkaian pembelajaran mesin, bahasa dan teknik statistik, yang biasanya digunakan untuk memodelkan dan mengekstrak informasi dari teks utama untuk tujuan analisis bisnis dan tujuan lainnya[8].

2.4 Text Preprocessing

Text preprocessing dilakukan untuk menghilangkan *noise* pada data seperti simbol, tanda baca, singkatan, ejaan yang salah, kata-kata yang tidak berarti dan lain sebagainya untuk mendapatkan informasi yang berguna[9]. *Text preprocessing* biasanya terdiri dari beberapa tahap yaitu *cleaning*, *case folding*, *tokenizing*, *stemming*, dan *filtering*[9]. *Text preprocessing* juga dapat dilakukan untuk mengatasi *noise* secara spesifik seperti melakukan normalisasi Bahasa[9]. Dalam penelitian ini tahapan *stemming* tidak dilakukan karena *stemming* dapat menyebabkan menurunnya nilai akurasi dari model[10].

a) *Cleaning*

Proses *cleaning* berguna untuk membersihkan data dari karakter-karakter yang dapat mengganggu proses selanjutnya seperti angka, tanda baca, URL, nama pengguna, dan lain sebagainya.

b) *Case Folding*

Proses *case folding* bertujuan untuk menyamaratakan bentuk setiap huruf yang ada ke dalam satu format yang sama yaitu *uppercase* (huruf besar) atau *lowercase* (huruf kecil).

c) *Tokenizing*

Proses ini memecah setiap dokumen yang ada ke dalam bentuk kata-kata yang Menyusun dokumen tersebut yang disebut dengan *token*.

d) *Normalisasi Bahasa*

Normalisasi Bahasa mengubah kata-kata yang tidak berada pada bentuk baku atau sesuai dengan KBBI seperti kata-kata *slang* (gaul) ke dalam bentuk baku nya, seperti kata “gk” dan “tdk” akan diubah menjadi “tidak”.

e) **Filtering**

Proses *filtering* menghilangkan kata-kata yang umum digunakan atau kata-kata yang tidak berarti sehingga kata-kata yang tersisa adalah kata-kata yang memiliki makna penting.

2.5 Pembobotan Term dengan TF-IDF

Pembobotan kata adalah tahapan yang dilakukan untuk menentukan seberapa penting kata tersebut dalam dokumen, dimana semakin tinggi nilainya maka kata tersebut semakin penting begitu pula sebaliknya[11]. Salah satu cara untuk menghitung bobot kata adalah dengan menggunakan *Term Frequency – Inverse Document Frequency* (TF-IDF) dimana cara ini sangat umum digunakan karena mudah dipakai dan hasilnya akurat serta efisien.

$$W_{t,d} = tf_{t,d} \times \log\left(\frac{N}{df_t}\right) \quad (2)$$

Keterangan dari rumus *TF – IDF* di atas adalah sebagai berikut

wt,d : Bobot kata t pada dokumen d
 tft,d : frekuensi kata t pada dokumen d
 N : Jumlah seluruh dokumen
 dft : Jumlah dokumen yang mengandung kata t

2.6 Naïve Bayes (NB)

Naive Bayes adalah salah satu metode klasifikasi yang mana cara kerja metode ini berdasarkan pada peluang/probabilitas dan Teorema Bayes, metode ini berasumsi bahwa ada atau tidaknya suatu atribut tidak memiliki hubungan dengan atribut lainnya[12]. Metode klasifikasi ini bersifat *supervised learning* dimana terdapat peranan manusia sebagai *supervisor* dalam proses pembelajarannya. Performa dari *naive bayes* ini memiliki waktu yang singkat sehingga proses analisis sentiment dapat dilakukan dengan waktu yang singkat[13].

Naive Bayes memiliki beberapa jenis, yaitu *Gaussian*, *Multinomial*, dan *Bernoulli*, Dalam penelitian ini jenis *naive bayes* yang dipakai adalah *multinomial naive bayes* karena memiliki perhitungan yang sederhana dan tingkat akurasi yang tinggi[14].

$$C_{map} = \arg \max_{c \in V} P(p) \prod_{t=1}^{|V|} P(W_i|p) \quad (2)$$

Keterangan:

$P(p)$: Peluang kemunculan dokumen yang berada pada kelas p.
 $P(W_i|p)$: Peluang kemunculan W_i pada kelas p.

Untuk mendapatkan nilai peluang kemunculan dokumen pada suatu kelas, dapat dicari dengan menggunakan persamaan berikut :

$$P(p) = \frac{|doc p|}{|document|} \quad (3)$$

Keterangan:

$P(p)$: Peluang kemunculan dokumen pada tiap kelas p.
 $|doc p|$: Jumlah dokumen untuk tiap kelas p.
 $|document|$: Jumlah dokumen keseluruhan

Sementara untuk mendapatkan nilai kemunculan W_i pada kelas p dapat didapatkan dengan menggunakan persamaan berikut :

$$P(W_i|p) = \frac{count(W_{ij,p})}{|c|+|V|} \quad (4)$$

Keterangan:

$P(W_i|p)$: Peluang kemunculan kata W_i pada kelas p
 $Count(W_i,p)$: Jumlah kemunculan kata W_i pada kelas p

$|c|$: Total keseluruhan kata pada kelas p
 $|V|$: Total keseluruhan *term* (kata).

Untuk mencegah nilai probabilitas 0 (nol) pada kata yang tidak terjadi, maka dibutuhkan *laplace smoothing* untuk mengatasi hal tersebut[15]. Sehingga persamaan setelah ditambahkan *laplace smoothing* adalah sebagai berikut :

$$P(W_i|p) = \frac{\text{count}(w_{ij,p})+1}{|c|+|V|*1} = \frac{\text{count}(w_{ij,p})+1}{|c|+|V|} \quad (5)$$

2.7 Evaluasi

Untuk mengetahui seberapa baik kinerja algoritma *naïve bayes* yang digunakan dalam penelitian ini, digunakan *confusion matrix* yang akan memvisualisasikan kinerja dari algoritma. Biasanya, akurasi digunakan untuk mengukur kinerja algoritma, namun jika data yang dipakai tidak berimbang maka nilai akurasi tidak bisa lagi dijadikan sebagai acuan tunggal[16].

Tabel 1. *Confusion Matrix*

Aktual	Prediksi	
	Negatif	Positif
Negatif	TN	FP
Positif	FN	TP

- *True Positive* (TP): Jumlah data positif yang terdeteksi benar.
- *True Negative* (TN): Jumlah data negatif yang terdeteksi benar.
- *False Positive* (FP): Data negatif tetapi terdeteksi sebagai data positif.
- *False Negative* (FN): Data positif tetapi terdeteksi sebagai data negatif.

a) Akurasi : Persentase dari jumlah seluruh prediksi benar dibagi jumlah keseluruhan data.

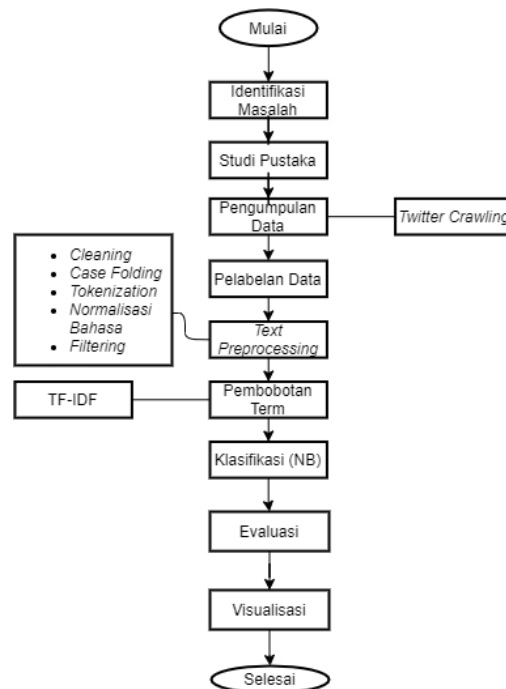
$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

b) *Specificity* : Persentase dari nilai yang salah berdasarkan fakta dibagi dengan jumlah keseluruhan data yang diprediksi salah[17].

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (7)$$

c) *Recall* : Persentase dari nilai yang benar berdasarkan fakta dibagi dengan jumlah keseluruhan data yang diprediksi benar.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (8)$$



Gambar. 1. Flowchart Penelitian

Penelitian dilakukan melalui beberapa tahap, tahap pertama adalah mengidentifikasi masalah lalu melakukan studi Pustaka untuk menjawab masalah tersebut. Selanjutnya adalah mengumpulkan data yang sesuai dengan topik penelitian dalam hal ini adalah data *tweet* yang berkaitan dengan Undang-Undang Cipta Kerja. Lalu data akan diberikan label oleh tiga *annotator* yang akan memberikan label pada tiap *tweet* yang sudah dikumpulkan. Setelah itu data yang sudah diberi label akan *preprocessing* terlebih dahulu untuk menghilangkan *noise* yang ada pada data, lalu dilakukan pembobotan tiap kata menggunakan TF-IDF. Setelah nilai bobot tiap kata berhasil didapatkan, selanjutnya adalah melakukan klasifikasi dengan menggunakan algoritma *naïve bayes* yang kemudian akan dilihat performanya apakah cukup baik atau tidak dalam melakukan klasifikasi terhadap sentimen positif dan negatif pada Undang-Undang Cipta Kerja.

3 Hasil dan Pembahasan

Penelitian ini menggunakan data yang berasal dari media social *twitter* dengan menggunakan kata kunci untuk pencarian yaitu #UUCIPTAKERJA yang diambil dari 5 Oktober – 1 November 2020 ketika UU Cipta Kerja disahkan, dan tanggal 2 – 30 November 2020 ketika UU Cipta Kerja resmi diberlakukan dan bertepatan dengan 1 tahun kepemimpinan Jokowi-Ma'ruf Amin. Data yang terkumpul berjumlah 1527 dokumen *tweet* pada tanggal 5 Oktober – 1 November 2020, dan 410 data pada tanggal 2 – 30 November 2020 dengan total dokumen *tweet* berjumlah 1937.

Data yang sudah dikumpulkan akan diberikan label secara manual oleh 3 *annotator*, hasil pelabelan dapat dilihat pada Tabel 2 sebagai berikut :

Tabel 2. Pelabelan Data

Data	Pihak 1	Pihak 2	Pihak3	Hasil Akhir
UUCiptaKerja memberikan ruang untuk UMKM berkembang sehingga kualitas pengusaha meningkat dan investor semakin berminat berinvestasi di Indonesia	Positif	Positif	Positif	Positif
@Lini_ZQ Kirain blunder presiden cukup di #OmnibusLaw taunya	Negatif	Negatif	Negatif	Negatif

malah merembet ke hewan purba.

Apabila dalam pelabelan ditemukan perbedaan pendapat, maka label yang digunakan adalah label yang paling banyak dipilih diantara *annotator*[18]. Setelah diberi label, data akan melalui *preprocessing* untuk menghilangkan *noise* yang ada pada data. Hasil *preprocessing* dapat dilihat pada Tabel 3.

Tabel 3. *Preprocessing* Data Sampel

Data sebelum preprocessing	Data sesudah preprocessing	Label
Walau pada ribut soal Omnibuslaw, kita harus tetap optimis untuk Indonesia lebih maju, selalu taati juga protokol kesehatan #sumpahpemuda #sumpahpemuda5.0#epivideopedia#covid19#kesehatanpulih#ekonomibangkit #UUCiptaKerja #omnibuslaw #epicentrumnews https://t.co/zg35MOXgCN	ribut omnibuslaw optimis maju taati protokol kesehatan	Positif
UU Omnibuslaw Cipta Kerja Dukungan kemudahan bagi para pelaku UMKM #OmnibusLaw	undang undang cipta kerja dukungan kemudahan pelaku umkm	Positif
UU CILAKA/#OmnibusLaw akan sama sperti UU KPK,... apapun caranya rakyat menolak, mau demo sperti apapun tetap aja Penguasa tidak akan peduli..mrk akan tetap melanjutkan apa yg mereka mau. Siapa yg bersuara keras protes akan ditangkap. UU KPK saja bedhasil apalagi UU skrg ini.	undang undang celaka undang undang kpk apapun rakyat menolak demo apapun penguasa peduli melanjutkan bersuara keras protes ditangkap undang undang kpk berhasil undang undang	Negatif
OmnibusLaw #UUCiptaKerja menambah kesejahteraan Buruh https://t.co/xpimgSOJITC	menambah kesejahteraan buruh	Positif

Setelah *preprocessing* selesai dilakukan, terdapat dokumen yang kosong dan duplikat sehingga harus dihilangkan dari data, jumlah keseluruhan dokumen dari jumlah awal sebanyak 1937 dokumen, berubah menjadi 1871 data. Selanjutnya adalah melakukan pembobotan TF-IDF terhadap setiap kata yang ada pada *tweet* dari data yang digunakan pada penelitian ini sejumlah 1871 data menggunakan persamaan (1).

Tabel 4 Perhitungan TF-IDF

Term	Dokumen				DF	IDF	TF-IDF			
	D1	D2	D3	D4			D1	D2	D3	D4
apapun	0	0	2	0	1	0,602	0	0	1,204	0
berhasil	0	0	1	0	1	0,602	0	0	0,602	0
bersuara	0	0	1	0	1	0,602	0	0	0,602	0
buruh	0	0	0	1	0	0,602	0	0	0	0,602
buruh	0	1	0	0	0	0,602	0	0,602	0	0
celaka	0	1	0	0	0	0,602	0	0,602	0	0
cipta	0	0	1	0	0	0,602	0	0	0,602	0
demo	0	0	1	0	0	0,602	0	0	0,602	0
ditangkap	0	1	0	0	0	0,602	0	0,602	0	0
dukungan	0	1	0	0	0	0,602	0	0,602	0	0
kemudahan	0	0	1	0	0	0,602	0	0	0,602	0
keras	0	1	0	0	0	0,602	0	0,602	0	0
kerja	1	0	0	0	1	0,602	0,602	0	0	0
kesehatan	0	0	0	1	0	0,602	0	0	0	0,602
kesejahteraan	0	0	2	0	0	0,602	0	0	1,204	0
kpk	1	0	0	0	1	0,602	0,602	0	0	0

Dari 1871 data yang digunakan pada penelitian ini, setelah dilakukan pembobotan kata menggunakan TF-IDF

menghasilkan 4601 variabel kata dengan nilai bobotnya masing-masing. Setelah didapatkan nilai bobot dari setiap kata, selanjutnya adalah melakukan proses klasifikasi menggunakan *naïve bayes*. Data sebanyak 1871 akan dibagi menjadi data latih dan data uji menggunakan rasio 80:20, rasio 80:20 digunakan karena berdasarkan [19] 80:20 merupakan tolak ukur yang baik dalam masalah data *imbalance*. Data yang digunakan sebagai data latih untuk membuat model berjumlah 80% dari 1871 yaitu 1496 data, dan untuk pengujian menggunakan data uji sebanyak 20% dari 1871 yaitu 375 data. Rincian jumlah data yang digunakan dapat dilihat pada Tabel 5.

Tabel 5 Pembagian Data

	Label Positif	Label Negatif	Total Jumlah
Data Latih	917	579	1496
Data Uji	230	145	375
Total	1147	724	1871

Dengan menggunakan *multinomial naïve bayes* didapatkan hasil klasifikasi yang dapat dilihat pada Tabel 6 berupa *confussion matrix* yang akan digunakan untuk menghitung performa dari algoritma tersebut.

Tabel 6 *Confussion Matrix* Hasil Klasifikasi

Aktual	Prediksi	
	Negatif	Positif
Negatif	51(TN)	94 (FP)
Positif	12(FN)	218(TP)

Dari hasil pada Tabel 6 diatas, menggunakan persamaan (6), (7), dan (8), performa model *naïve bayes* mendapatkan nilai 71,73% akurasi, 94,7% *specificity*, dan 35,17% *recall*, serta waktu eksekusi selama 0,09984 detik. Dari hasil performa diatas terlihat bahwa nilai *recall* dan *specificity* terpaut sangat jauh, hal ini disebabkan karena jumlah data antar kelas pada data latih tidak berimbang. Oleh karena itu diperlukan penyeimbangan jumlah data pada data latih, Dalam penelitian ini, untuk menyeimbangkan jumlah data antar kelas pada data latih digunakan metode *random undersampling*. *Random Undersampling* akan menghapus secara acak data pada kelas yang memiliki jumlah data lebih banyak, hingga jumlahnya sama dengan jumlah data pada kelas dengan jumlah data yang lebih sedikit. *Random Undersampling* dipilih karena dapat mempercepat waktu eksekusi dan meningkatkan performa algoritma[20].

Hasil klasifikasi model setelah dilakukan *undersampling* dapat dilihat pada Tabel 7 sebagai berikut :

Tabel 7 *Confussion Matrix* Hasil Klasifikasi dengan *Undersampling*

Aktual	Prediksi	
	Negatif	Positif
Negatif	107(TN)	38(FP)
Positif	35(FN)	195(TP)

Performa model yang didapatkan setelah melakukan *undersampling* dengan menggunakan persamaan (6),(7), dan (8) dibandingkan dengan performa model sebelumnya dengan data *unbalanced* memiliki hasil sebagai berikut: nilai akurasi sebesar 80,53% yang mana pada data *unbalanced* hanya sebesar 71,73%, *specificity* sebesar 73,79% yang mana pada data *unbalanced* hanya 35,17%, dan *recall* sebesar 84,78% yang mana pada data *unbalanced* sebesar 94,7%, serta waktu eksekusi selama 0,07679 detik yang mana lebih cepat dari model dengan data *unbalanced* yaitu selama 0,09984 detik. Hal ini menunjukkan bahwa performa model untuk mengklasifikasikan sentimen positif dan negatif cukup baik.

Dari data yang ada kemudian dilakukan visualisasi untuk mengetahui *wordcloud* yang menunjukkan kata positif dan negatif pada data setelah pengesahan UU Cipta Kerja dan setelah pengundangan UU Cipta Kerja. Hasil Visualisasi pada data setelah pengesahan UU Cipta Kerja adalah sebagai berikut.

- Sentiment Analysis* (2nd ed.). Springer. <http://www.springer.com/series/13199>
- [6] Liu, B. (2012). Opinion spam detection. In *Sentiment Analysis and Opinion Mining* (Issue May). Morgan & Claypool Publishers. https://doi.org/10.1142/9789813100459_0007
- [7] Aggarwal, C. C., & Zhai, C. X. (2013). Mining text data. In *Mining Text Data* (Vol. 9781461432234). <https://doi.org/10.1007/978-1-4614-3223-4>
- [8] Sarkar, D. (2016). Text Analytics with Python: A Practical Real-World Approach to Gaining Actionable Insights from Your Data. In *European Physical Journal A* (Vol. 32, Issue 1). Apress. <https://doi.org/10.1140/epja/i2006-10279-1>
- [9] Hidayatullah, A. F., & Ma'arif, M. R. (2017). Pre-processing Tasks in Indonesian Twitter Messages. *Journal of Physics: Conference Series*, 801(1). <https://doi.org/10.1088/1742-6596/755/1/011001>
- [10] 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), 127–132. <https://doi.org/10.11591/eecsi.v2i1.791>
- [11] Melita, R., Amrizal, V., Suseno, H. B., & Dirjam, T. (2018). Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim). *Jurnal Teknik Informatika*, 11(2), 149–164. <https://doi.org/10.15408/jti.v11i2.8623>
- [12] Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *INTEGER: Journal of Information Technology*, 1(1), 32–41. https://www.researchgate.net/profile/Ghulam_Buntoro/publication/316617194_Analisis_Sentimen_Calon_Gubernur_DKI_Jakarta_2017_Di_Twitter/links/5907eee44585152d2e9ff992/Analisis-Sentimen-Calon-Gubernur-DKI-Jakarta-2017-Di-Twitter.pdf
- [13] Gunawan, B., Pratiwi, H. S., & Pratama, E. E. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 4(2), 113. <https://doi.org/10.26418/jp.v4i2.27526>
- [14] Sabrani, A., Wedashwara W., I. G. W., & Bimantoro, F. (2020). Multinomial Naïve Bayes untuk Klasifikasi Artikel Online tentang Gempa di Indonesia. *Jurnal Teknologi Informasi, Komputer, Dan Aplikasinya (JTika)*, 2(1), 89–100. <https://doi.org/10.29303/jtika.v2i1.87>
- [15] Bunga, M. T. H., S, B., Djahi, & Nabuasa, Y. Y. (2018). Multinomial Naive Bayes Untuk Klasifikasi Status Kredit Mitra Binaan Di Pt . Angkasa Pura I Program Kemitraan. *J-Icon*, 6(2), 30–34. <https://media.neliti.com/media/publications/292504-multinomial-naive-bayes-untuk-klasifikas-2536567f.pdf>
- [16] Jurafsky, D., & Martin, J. H. (2020). *Speech and Language Processing: An introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (3rd ed.). <http://www.cs.colorado.edu/~martin/slp.html>
- [17] Habib, A., Alalyani, M., Hussain, I., & Almutheibi, M. S. (2015). Brief review on Sensitivity, Specificity and Predictivities. *IOSR Journal of Dental and Medical Sciences*, 14(4), 2279–2861. <https://doi.org/10.9790/0853-14456468>
- [18] Rogers, A., Romanov, A., Rumshisky, A., Volkova, S., Gronas, M., & Gribov, A. (2018). RuSentiment: an enriched sentiment analysis dataset for social media in Russian. *Proceedings of the 27th International Conference on Computational Linguistics*, 755–763. <http://www.aclweb.org/anthology/C18-1064>
- [19] Koch, R. (2013). The 80/20 Principle And 92 Other Powerful Laws Of Nature. In *Long Range Planning*. Quercus. [https://doi.org/10.1016/s0024-6301\(97\)80978-8](https://doi.org/10.1016/s0024-6301(97)80978-8)
- [20] Irawan, E., & Wahono, R. S. (2015). Penggunaan Random Under Sampling untuk Penanganan Ketidakseimbangan Kelas pada Prediksi Cacat Software Berbasis Neural Network. *Journal of Software Engineering*, 1(2), 92–100.