

ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP KEBIJAKAN SISTEM ZONASI PPDB MENGGUNAKAN ALGORITMA MULTINOMIAL NAIVE BAYES

Lani Lesmana¹, Mukrodin², Fuaida Nabyla³

¹Universitas Peradaban

²Universitas Peradaban

³Universitas Peradaban

Email: ¹lany746842@yahoo.com, ²mukrodins@gmail.com, ³fnabyla@gmail.com

Abstrak

Penerimaan Peserta Didik Baru (PPDB) merupakan suatu kegiatan tahunan yang diselenggarakan oleh panitia tingkat sekolah di bawah pengawasan dan koordinasi dinas pendidikan, dimana kegiatan ini merupakan tahap seleksi bagi calon peserta didik baru. Kebijakan PPDB dengan sistem zonasi telah diberlakukan sejak tahun 2017 yang diharapkan dapat memberikan layanan pendidikan yang bermutu secara merata bagi warga anggota masyarakat. Namun dalam pelaksanaannya mengalami kendala yang seringkali merugikan masyarakat. Hal ini menjadikan kebijakan yang dilaksanakan selalu menuai konflik dan protes dari masyarakat. Tidak sedikit dari masyarakat yang mengekspresikan tanggapan-tanggapan mengenai sistem zonasi ini pada sosial media *Twitter*. Opini masyarakat yang terdapat di *tweet* tersebut dapat dimanfaatkan sebagai bahan analisis sentimen untuk mengetahui masyarakat terhadap kebijakan sistem zonasi apakah positif atau negatif. *Multinomial Naive Bayes* merupakan salah satu teknik *Machine Learning* yang populer untuk klasifikasi teks, karena sangat sederhana, efisien dan memiliki performa yang baik. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap kebijakan sistem zonasi dengan menerapkan metode klasifikasi atau algoritma *Machine Learning* yaitu *Multinomial Naive Bayes*. Hasil dari penelitian ini diharapkan dapat membantu mengevaluasi kebijakan yang ada, diharapkan hal-hal yang sudah baik agar tetap dipertahankan dan hal-hal yang buruk dapat diperbaiki agar kebijakan yang ada berjalan sesuai harapan.

Keyword: PPDB, sistem zonasi, analisis sentimen, *twitter*, *multinomial naive bayes*

I. PENDAHULUAN

Pelayanan pendidikan yang merata dan berkeadilan bagi setiap warga negara merupakan suatu kewajiban atau keharusan pemerintah yang dijamin Undang-Undang Dasar 1945. Namun sekarang ini banyak ditemukannya sekolah-sekolah favorit atau unggulan, dimana sekolah-sekolah ini terkesan diberikannya perlakuan yang berbeda. Dengan adanya sekolah unggulan atau favorit orientasi pembinaan dan perlakuan khusus mengakibatkan semua sumber daya diberikan kepada sekolah tersebut. Disisi lain terdapat sekolah reguler yang kurang berkembang, hal ini terjadi karena kurangnya perhatian terhadap sekolah reguler, sementara sekolah unggulan yang dijadikan pusat perhatian dan terus diberi bantuan dan dukungan yang berlebihan. Akibatnya, penyebaran mutu sekolah juga tidak bisa dilakukan oleh pemerintah yang mengakibatkan mutu layanan pendidikan tidak bisa merata dan adil untuk melayani anak di setiap wilayah pelosok tanah air.

Dengan adanya kebijakan zonasi ini tentu memiliki kendala dalam pelaksanaannya. Salah satu masalah yang dihadapi yaitu pihak sekolah yang tidak siap dalam menerima sistem tersebut, dan juga kurangnya sosialisai pemerintah

terhadap masyarakat[1]. Hal ini menjadikan kebijakan yang dilaksanakan selalu menuai konflik dan protes dari masyarakat yang merasa dirugikan. Hal ini terjadi karena kebijakan yang dibuat tidak sesuai dengan perbedaan kondisi serta kebutuhan dari setiap daerah di Indonesia (Azanella, 2019). Tidak sedikit Masyarakat yang menyampaikan opini mereka melalui sosial media. Opini-opini yang ada tidak hanya komentar yang bersifat positif melainkan juga komentar yang bersifat negatif, yang merepresentasikan kepuasan saat menjalankan kebijakan ini.

Tanggapan-tanggapan yang dilontarkan oleh pengguna *Twitter* mengenai kebijakan sistem zonasi mengandung sentimen yang dapat dianalisis. Analisis sentimen merupakan bidang studi yang mempelajari tentang menganalisis pendapat, evaluasi, sikap, dan emosi seseorang dari bahasa tulisan dan salah satu bidang penelitian yang paling aktif dari bahasa alami [2]. Analisis sentimen akan mengklasifikasi kalimat untuk mengetahui pendapat yang dikemukakan dalam kalimat apakah bersifat positif atau negatif.

Salah satu metode yang dapat digunakan dalam klasifikasi adalah *Multinomial Naive Bayes*. *Multinomial Naive Bayes* dapat dikatakan sebagai versi upgrade dari klasifikasi *Naive Bayes* yang ada, dan secara efektif memanipulasi jumlah kata dengan menghitung frekuensi

setiap kata, [3]. Metode *Multinomial Naive Bayes* mengikuti prinsip distribusi *Multinomial* yang digunakan untuk pemrosesan teks, Selain itu *Multinomial Naive Bayes* digunakan karena kecepatan dan kesederhanaannya dalam klasifikasi teks [4].

II. TINJAUAN PUSTAKA

A. Studi Literatur

Penelitian yang dilakukan oleh sudiantoro dan zuliarso mengenai analisis sentimen pada media sosial *twitter* berupa *tweet*, dimanfaatkan sebagai bahan analisis sentimen untuk mengetahui masyarakat terhadap pilkada jawa barat apakah positif atau negatif [5]. Data yang digunakan terdiri dari 300 data *tweet* yang dibagi menjadi 2 yaitu untuk data latih sebanyak 200 data dan 100 data untuk data uji. Sebelum dilakukan klasifikasi, peneliti melakukan beberapa tahap pemrosesan teks seperti, normalisasi, *case folding*, *stopwords removal* dan *tokenisasi*. Klasifikasi yang digunakan adalah *naive bayes*.

Penelitian yang dilakukan oleh purwanto dan santoso mengenai analisis sentimen digunakan untuk menentukan review yang diberikan oleh setiap produk tersebut positif atau negatif [6]. Data yang digunakan dalam penelitian ini diambil dari *website* penjualan *qoo10.co.id*. Review yang didapat dengan menggunakan *web crawler*, dilakukan preproses yaitu dengan melakukan penghilangan *alphanumeric*, *case folding*, dan *stopword removal*. Gambar 2.1 menjelaskan secara singkat mengenai penelitian purwanto dan santoso.

Penelitian yang dilakukan oleh Rahman, wiranto dan Afrizal mengenai klasifikasi teks menggunakan algoritma *Multinomial Naive Bayes*, penelitian ini bertujuan untuk mengkalsifikasi berita secara otomatis [7]. Dalam penelitian ini data diperoleh dari lima media online yaitu *detik.com*, *viva.co.id*, *inilah.com*, *antaranews.com*, dan *okezone.com*. Jumlah data berita yang digunakan sebanyak 1011 data dengan total isu yang dikelompokkan sebanyak 15 kelas. Klasifikasi yang digunakan dalam penelitian ini yaitu *Naive Bayes* dengan model *Multinomial* dan seleksi fitur menggunakan *Document Frequency Thresholding* serta pembobotan dengan menggunakan *Term Frequency-Inverse*.

B. Twitter

Twitter merupakan sebuah situs layanan *Microblogging*, yang penggunaannya dapat mengirim dan membaca status terbaru seseorang [12]. Status terbaru dalam *Twitter* disebut dengan *tweets*. *Twitter* dikategorikan sebagai situs *Microblogging* karena update (*tweets*) dibatasi dengan 140 karakter saja [13].

C. Analisis Sentimen

Analisis sentimen adalah pengklasifikasian sentimen positif atau negatif dari hasil ekstraksi pendapat seseorang terhadap sesuatu hal. Secara garis besar analisis sentimen digunakan untuk memberikan pandangan terhadap suatu produk, merk, reputasi, reviews, blog, diskusi, berita, komen, dan *feedback* [8]. Analisis sentimen dapat digunakan untuk

mengungkap opini publik terhadap suatu isu yang ada didalam masyarakat.

Analisis sentimen atau penambangan opini adalah studi komputasi atas opini, penilaian, sikap, dan emosi orang terhadap entitas, individu, masalah, peristiwa, topik, dan atributnya [9]. Dengan pertumbuhan media sosial (seperti, ulasan, forum diskusi, blog, *micro-blog*, dan jejaring sosial) penelitian mengenai penambangan teks banyak dimanfaatkan bahkan diluar ilmu komputer, seperti pemanfaatan dalam ilmu sosial dan ilmu manajemen karena pentingnya untuk kebutuhan bisnis dan masyarakat secara keseluruhan.

D. Multinomial Naive Bayes

Multinomial Naive Bayes merupakan salah satu metode spesifik dari metode *Naive Bayes*. *Multinomial Naive Bayes* ini juga merupakan salah satu *machine learning* dalam *supervised learning* pada proses pengklasifikasian teks dengan menggunakan nilai probabilitas suatu kelas dalam suatu dokumen [10]. *Naive Bayes* bekerja berdasarkan probabilitas kondisional (mempertimbangkan independensi kondisional fitur), sedangkan *Multinomial Naive Bayes* bekerja berdasarkan distribusi *multinomial*. Pengelompokan *Multinomial Naive Bayes* mempertimbangkan banyak kejadian dari setiap istilah [3].

Proses pembelajaran dimulai dengan memasukkan data latih yang digunakan untuk pembelajaran. Kemudian mulai untuk pembentukan daftar kata-kata. Daftar kata-kata merupakan kumpulan kata unik yang berasal dari data latih. Kemudian menghitung *prior* atau peluang kemunculan suatu kategori pada semua dokumen latih dapat dilakukan dengan menggunakan persamaan berikut [11]:

$$P(c) = \frac{N_c}{N_{doc}} \quad (1)$$

Keterangan:

- N_c : Banyaknya kategori c pada dokumen latih
- N_{doc} : Banyaknya keseluruhan dokumen latih yang digunakan.

Dilanjutkan dengan menghitung peluang sebuah kata i masuk ke dalam kategori atau kelas tertentu dapat dilakukan dengan menggunakan persamaan berikut [11]:

$$P(w_i, c) = \frac{\text{count}(w_i, c) + 1}{(\sum_w \text{count}(w, c)) + V} \quad (2)$$

$\text{Count}(w_i, c)$ merupakan jumlah kata tertentu yang muncul dalam suatu kategori atau kelas. Penambahan nilai satu berfungsi agar tidak menghasilkan probabilitas bernilai 0, modifikasi ini disebut dengan *smoothing*. Terdapat *laplace smoothing* jika nilai *smoothing* adalah satu dan terdapat *add-k smoothing* jika nilai *smoothing* tidak bernilai satu. $\sum_w V \text{count}(w, c)$ merupakan jumlah seluruh kata pada kelas. Sedangkan $|V|$ merupakan jumlah seluruh kata unik diseluruh kelas. Setelah tahap pembelajaran selesai dilakukan maka tahap selanjutnya adalah tahap klasifikasi data baru berdasarkan hasil pembelajaran. Berikut ini merupakan rumus yang digunakan dalam melakukan klasifikasi data baru [11]:

$$C_{NB} = \underset{c \in C}{\operatorname{argmax}} \log P(c) + \sum_{i \in \text{positions}} \log P(w_i | c) \quad (3)$$

III. METODE PENELITIAN

Metode yang digunakan pada penelitian ini adalah penelitian eksperimen dengan tahapan penelitian sebagai berikut:

1. Pengumpulan Data

Data *tweet* mengenai sistem zonasi diambil dari sosial media *Twitter*. Data yang digunakan sebanyak 3000 komentar. Data yang ada disimpan dengan format file csv.

2. Deteksi *Bot Spammer*

Karakteristik dari *tweet* yang dikategorikan spam memiliki sentimen emosi netral atau tidak ada preferensi tertentu terhadap suatu perspektif dari user yang memposting *tweet*. Oleh karena itu perlu dihilangkan untuk mendapatkan sentimen yang sesuai. Seleksi secara manual berdasarkan karakteristik tertentu yang menandakan *tweet* tersebut merupakan spam.

3. *Preprocessing*

Proses *Preprocessing* meliputi proses *Case Folding*, *Cleaning*, *Negation Handling*, dan *Stopword Removal*. Proses ini berfungsi untuk mengurangi atribut yang dianggap tidak berpengaruh dan mengurangi dimensi kata.

4. *Labelling*

Pada tahap *Labelling* dataset, *Labelling* dilakukan menggunakan pendekatan *Lexicon* dengan menggunakan kamus positif dan kamus negatif. Pendekatan ini dilakukan untuk menentukan apakah data *tweet* bersentimen positif atau negatif sesuai kamus yang ada. Pada proses *Labelling* kata pada data *tweet* akan dicocokkan dengan kata yang berada pada kamus kata positif dan negatif.

5. *Modelling*

Proses *Modelling* merupakan proses membuat model klasifikasi *Multinomial Naive Bayes*. Terdapat beberapa tahapan yang dilakukan seperti pemilihan fitur, *training* model, dan *evaluasi*.

IV. HASIL PENELITIAN DAN PEMBAHASAN

1. Pengumpulan Data

Data yang dikumpulkan bersumber dari media sosial *Twitter* yang diambil dari jangka waktu 01 Januari 2018 sampai dengan 06 Mei 2020. Data *tweet* yang digunakan merupakan data *tweet* yang berhubungan dengan Sistem Zonasi PPDB. Pengumpulan data *tweet* dilakukan menggunakan fungsi python yaitu *Twitterscraper*. Jumlah data yang didapatkan sejumlah 3000 komentar.

2. *Bot Spammer*

Pelabelan pada *tweet* yang dianggap berasal dari bot/spam dilakukan dengan cara manual. *Tweet* yang dikategorikan spam kemudian dipisahkan karena memiliki sentimen emosi

netral atau tidak ada preferensi tertentu terhadap suatu perspektif dari user yang memposting *tweet*. Dari hasil pelabelan yang dilakukan terdapat 477 *tweet* yang dikategorikan spam.

3. *Preprocessing Data*

Tahapan *Preprocessing* data terdiri dari *Case Folding*, *Cleaning*, *Negation Handling*, dan *Stopword Removal*.

a) *Case Folding*

Tweet dari data set akan dilakukan proses *Case Folding* yaitu, proses mengubah semua huruf menjadi huruf kecil. Hal ini dilakukan agar data *tweet* yang ada konsisten menggunakan huruf kecil.

b) *Cleaning*

Pada proses *Cleaning* ini, semua kata yang ada di dalam tiap dokumen dikumpulkan dan dihilangkan tanda bacanya, serta dihilangkan jika terdapat simbol atau apapun yang bukan huruf. Hal ini dilakukan untuk menghapus karakter-karakter yang tidak digunakan dalam analisis sentimen.

c) *Negation Handling*

Pada tahap ini program akan menggabungkan kata negasi “tidak” dan “belum” dengan kata yang muncul setelah kata negasi. Kata negasi akan dihubungkan dengan kata selanjutnya menjadi “tidak_” + kata dan “belum_” + kata.

d) *Stopword Removal*

Pada tahap ini program akan menghapus kata pada *tweet* yang terdapat pada kamus kata *stopword*. Kata-kata yang dihilangkan merupakan kata-kata yang tidak mempunyai makna tersendiri jika dipisahkan dengan kata yang lain dan tidak terkait dengan dengan kata sifat yang berhubungan dengan sentimen.

e) *Labelling Dataset*

Setelah proses *preprocessing* selesai dilakukan, maka tahap selanjutnya adalah proses *labelling*. Proses *labelling* menggunakan pendekatan *lexicon* dimana berisi kamus kata positif dan negatif. Hasil dari proses *labelling* digunakan sebagai pembuatan model klasifikasi.

4. *Modelling Klasifikasi*

Modelling dilakukan untuk membuat model klasifikasi baru dengan metode *Multinomial Naive Bayes*. Pada tahap ini data yang telah melalui tahap *preprocessing* akan diolah menggunakan aplikasi *Jupyter Notebook*. Pada penelitian ini akan dibuat 2 model untuk klasifikasi, yaitu model untuk klasifikasi dua kelas (Positif dan Negatif) dan model untuk klasifikasi 3 kelas (Positif, Negatif, dan Netral).

a. *Klasifikasi Dua Kelas (Positif dan Negatif)*

Pembagian yang telah dilakukan oleh program menghasilkan 978 data *training* dan 245 data *testing* yang dilakukan acak oleh sistem. Selanjutnya, Data yang ada agar dapat digunakan, maka perlu dilakukan konversi teks kedalam

bentuk numerik ataupun vektor. Hal ini dilakukan karena mesin tidak dapat membaca teks.

Berdasarkan hasil dari uji coba model Klasifikasi *Multinomial Naive Bayes* menggunakan 978 data *training* dan 245 data *testing* diperoleh hasil seperti pada tabel 4.1.

Tabel 4.1 Hasil Fitur Pada Model Klasifikasi Dua Kelas

Parameter	Jumlah Fitur	Test Akurasi
0	4364.0	0.779592
1	1481.0	0.804082
2	905.0	0.795918
3	624.0	0.795918
4	469.0	0.787755

Dari hasil tersebut *Test accuracy* atau akurasi uji coba terhadap data *testing* terbaik terjadi pada saat fitur yang digunakan berjumlah 1481 dengan frekuensi kemunculan fitur atau kata minimal 1 kali pada data *training* atau parameter ke-1. Nilai *test accuracy* sebesar 80%.

Evaluasi model dilakukan dengan melakukan uji coba melakukan prediksi terhadap data *testing*. Berdasarkan hasil dari uji coba model Klasifikasi *Multinomial Naive Bayes* terhadap data *testing* diperoleh *confusion matrix* seperti Tabel 4.2 berikut:

Tabel 4.2 *Confusion Matrix* Hasil Testing Dua Kelas

Hasil Labelling	Prediksi Negatif	Prediksi Positif
Negatif	61	32
Positif	18	134

Berdasarkan *confusion matrix* pada tabel 4.2 dapat diperoleh hasil perhitungan *accuracy*, *recall*, dan *precision* dengan perhitungan sebagai berikut:

a) Menghitung *Accuracy*

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100\%$$

$$Accuracy = \frac{61 + 134}{61 + 134 + 18 + 32} * 100\% = 80\%$$

b) Menghitung *Recall*

$$Recall\ positif = \frac{134}{134 + 18} * 100\% = 88\%$$

$$Recall\ negatif = \frac{61}{61 + 32} * 100\% = 66\%$$

$$Average\ Recall = \frac{88\% + 66\%}{2} = 77\%$$

c) Menghitung *Precision*

$$Precision\ positif = \frac{134}{134 + 32} * 100\% = 80\%$$

$$Precision\ negatif = \frac{61}{61 + 18} * 100\% = 77\%$$

$$Average\ Precision = \frac{80\% + 77\%}{2} = 79\%$$

Berdasarkan hasil perhitungan di atas dapat diketahui bahwa hasil uji coba menggunakan data *testing* menghasilkan tingkat *accuracy* 80%, *average recall* sebesar 77%, dan *average precision* sebesar 79%. Dengan akurasi lebih dari 70% maka model dapat dikatakan baik.

b. Klasifikasi Tiga Kelas (Positif, Negatif, dan Netral)

Pembagian yang telah dilakukan oleh program menghasilkan 2018 data *training* dan 505 data *testing* yang dilakukan acak oleh sistem. Sama dengan proses pada klasifikasi dua teks, pada klasifikasi ini juga perlu dilakukan Konversi teks menjadi bentuk numerik ataupun vektor. Berdasarkan hasil dari uji coba model Klasifikasi *Multinomial Naive Bayes* menggunakan 2018 data *training* dan 505 data *testing* diperoleh hasil seperti pada tabel 4.3

Tabel 4.3 Hasil Fitur Pada Model Klasifikasi Tiga Kelas

Parameter	Jumlah Fitur	Test Akurasi
0	6222.0	0.615842
1	2181.0	0.637624
2	1361.0	0.649505
3	993.0	0.629703
4	743.0	0.635644

Test accuracy atau akurasi uji coba terhadap data *testing* terbaik terjadi pada saat fitur yang digunakan berjumlah 1361 dengan frekuensi kemunculan fitur atau kata minimal 2 kali pada data *training* atau parameter ke-2. Nilai *test accuracy* sebesar 65%.

Evaluasi model dilakukan dengan melakukan uji coba melakukan prediksi terhadap data *testing*. Berdasarkan hasil dari uji coba model Klasifikasi *Multinomial Naive Bayes* terhadap data *testing* diperoleh *confusion matrix* seperti pada tabel 4.4 berikut:

Tabel 4.4 *Confusion Matrix* Hasil Testing Tiga Kelas

Hasil Labelling	Prediksi Negatif	Prediksi Positif	Prediksi Netral
Negatif	28	53	15
Positif	14	218	35
Netral	5	55	82

Berdasarkan *confusion matrix* pada tabel 4.4 dapat diperoleh hasil perhitungan *accuracy*, *recall*, dan *precision* dengan perhitungan sebagai berikut:

Tabel 4.15 Perhitungan *Confusion Matrix* Hasil Testing Tiga Kelas

Hasil Labelling	Prediksi Negatif	Prediksi Positif	Prediksi Netral	Total
Negatif	28	53	15	96
Positif	14	218	35	267
Netral	5	55	82	142
Total	47	326	132	505

a) Menghitung *Accuracy*

$$Accuracy = \frac{28 + 218 + 82}{505} * 100\% = 65\%$$

b) Menghitung *Recall*

$$Recall\ positif = \frac{82}{142} * 100\% = 58\%$$

$$Recall\ negatif = \frac{28}{96} * 100\% = 29\%$$

$$Recall\ netral = \frac{218}{267} * 100\% = 82\%$$

$$\text{Average Recall} = \frac{58\% + 96\% + 82\%}{3} = 56\%$$

c) Menghitung *Precision*

$$\text{Precision positif} = \frac{82}{132} * 100\% = 62\%$$

$$\text{Precision negatif} = \frac{28}{47} * 100\% = 62\%$$

$$\text{Precision netral} = \frac{218}{326} * 100\% = 67\%$$

$$\text{Average Precision} = \frac{62\% + 60\% + 67\%}{3} = 63\%$$

Berdasarkan hasil perhitungan di atas dapat diketahui bahwa hasil uji coba menggunakan data testing menghasilkan tingkat *accuracy* 65%, *average recall* sebesar 56%, dan *average precision* sebesar 63%. Dengan hasil akurasi kurang dari 70% dapat dikatakan model yang dibuat menghasilkan klasifikasi yang buruk.

V. KESIMPULAN

Pada penelitian ini dapat disimpulkan bahwa metode *Multinomial Naive Bayes* dapat digunakan untuk klasifikasi sentimen data tweet yang bersumber dari sosial media *Twitter*, terutama dalam mengklasifikasi sentimen bersifat positif dan negatif saja. Hal ini berdasarkan hasil uji coba model Klasifikasi menggunakan dua kelas (Positif dan Negatif) dan klasifikasi tiga kelas (Positif, Negatif, dan Netral), dimana pada klasifikasi dua kelas memiliki nilai akurasi mencapai 80%, sedangkan untuk klasifikasi tiga kelas (Positif, Negatif, dan Netral) menghasilkan akurasi yang buruk yaitu 65%. Hasil akurasi model *Multinomial Naive Bayes* juga dipengaruhi oleh fitur yang digunakan untuk melatih model, dapat dilihat pada tabel 4.1 dan tabel 4.3, dimana setiap perubahan fitur memiliki nilai akurasi yang berbeda.

VI. DAFTAR PUSTAKA

- [1] L. P. Wardhana and K. Supriyoko, "Manajemen Penerimaan Peserta Didik Baru Secara Online Berbasis Zonasi, Prestasi dan Perpindahan," *Media Manaj. Pendidik.*, vol. 2, no. 2, p. 228, 2019, doi: 10.30738/mmp.v2i2.5063.
- [2] B. Liu, *Sentiment Sentiment Analysis Analysis and and Opinion Opinion Mining Mining*. 2012.
- [3] P. P. M. Surya, L. V. Seetha, and B. Subbulakshmi, "Analysis of user emotions and opinion using Multinomial Naive Bayes Classifier," *Proc. 3rd Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2019*, pp. 410–415, 2019, doi: 10.1109/ICECA.2019.8822096.
- [4] A. A. Farisi, Y. Sibaroni, and S. Al Faraby, "Sentiment analysis on hotel reviews using Multinomial Naive Bayes classifier," *J. Phys. Conf. Ser.*, vol. 1192, no. 1, 2019, doi: 10.1088/1742-6596/1192/1/012024.
- [5] A. V. Sudiantoro, E. Zuliarso, and Studi, "ANALISIS

SENTIMEN TWITTER MENGGUNAKAN TEXT MINING DENGAN," pp. 398–401, 2018.

- [6] D. D. Purwanto and J. Santoso, "Multinomial Naive Bayes Classifier Untuk Menentukan Review," no. March, pp. 117–122, 2015, [Online]. Available: <https://www.researchgate.net/publication/319256329%0AMULTINOMIAL>.
- [7] A. Rahman, Wiranto, and D. Afrizal, "Online News Classification Using Multinomial Naive Bayes," vol. 6, no. 1, 2017.
- [8] V. Chandani, "Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film," 2015.
- [9] B. Liu and L. Zhang, *Chapter 13 A SURVEY OF OPINION MINING AND SENTIMENT ANALYSIS*. .
- [10] J. Wisudawati, Adiwijaya, and S. Al Faraby, "Klasifikasi Sentimen pada Movie Review dengan Metode Multinomial Naive Bayes Sentiment Classification Movie Review Using Multinomial Naive Bayes Method," vol. 4, no. 2, pp. 2978–2988, 2017.
- [11] D. Jurafsky and J. H. Martin, "Speech and Language Processing," 2019.
- [12] Adi, A. P., & Sanjaya, R. (2009). Panduan cepat menguasai twitter. Jakarta: PT Elex Media Komputindo.
- [13] Juju, D., & Studio, M. M. (2009). Twitter, tunggu apa lagi follow me! Jakarta: PT Elex Media Komputindo.