SIGMA
Jurnal Teknologi Pelita Bangsa

Vol. 11 No. 3 September 2020 ISSN: 2407-3903

Diterima, 16 Juli 2020 | Direvisi, 28 Agustus 2020 | Dipublikasikan, 28 September 2020

ANALISA SENTIMEN *TWEET* INDONESIA MENGGUNAKAN FITUR EKSTRASI DAN TEKNIK CROSS *VALIDATION* TERHADAP MODEL *NAÏVE BAYES*

Ahmad Turmudi[1], Khunaify Syarief Yasah[2]

Program Studi Teknik Informatika Fakultas Teknik Universitas Pelita Bangsa
[1]turmudi@pelitabangsa.ac.id

Abstraksi

Analisa sentimen merupakan ilmu di bidang studi pengolahan bahasa alami untuk menganalisa data berbentuk opini yang postif dan negatif dengan tujuan mendapakan hasil dalam pengambilan keputusan. Salah satu media dalam penelitian analisis sentimen ialah twitter. Masalah Utama dalam klasifikasi analisis sentimen adalah bagaimana memilih fitur dan validasi yang tepat dalam pengujian. Model yang digunakan untuk penelitian ini adalah Naïve Bayes. Naïve Bayes dapat dikombinasikan dengan *feature extraction*. Dalam pengujian ekstraksi fitur CountVectorizer dan *TFIDFVectorizer* dibandingkan dengan menggunakan teknik *Cross Validation* untuk memperbaiki klasifikasi *Naïve Bayes*. Pengukuran nilai dilakukan dengan membandingkan antara pengujian tanpa validasi dan menggunakan validasi. Akurasi dapat diukur dengan menggunakan confusion matrix, presisi dan recall. Hasil dari penelitian menunjukkan dengan menggunakan ekstraksi fitur *TF-IDFVectorizer* lebih baik dibandingkan *CountVectorizer* dengan mendapatkan akurasi tertinggi sebesar 85,98% dan untuk pengujian akhir fitur ekstrasi dengan *Cross Validation*lebih baik dibandingkan tidak menggunakan *Cross Validation* dengan mendapatkan nilai akurasi tertinggi sebesar 97,67%. Dengan demikian, pengujian fitur ekstrasi yang baik digunakan adalah *TF-IDFVectorizer* dan dengan menggunakan teknik *Cross Validation* dapat meningkatkan kinerja model *Naïve Bayes* dalam analisis sentiment*twitter* berbahasa Indonesia hingga mendapat selisih nilai akurasi sebesar 11,69%.

Kata Kunci : Analisis Sentimen, *twitter, Naïve Bayes, feature extraction, CountVectorizer, TF- IDFVectorizer, Crosss Validation.*

Abstract

Sentiment analysis is a science in the field of natural language processing studies to analyze data in the form of positive and negative opinions with the aim of getting results in decision making. One of the media in sentiment analysis research is twitter. The main problem in sentiment analysis classification is how to choosethe right features and validation in the test. The model used for this research is Naïve Bayes. Naïve Bayes canbe combined with feature extraction. In testing the feature extraction of CountVectorizer and TFIDFVectorizer is compared using the Cross Validation technique to improve the Naïve Bayes classification. Value measurement is done by comparing between testing without validation and using validation. Accuracy can be measured using confusion matrix, precision and recall. The results of the study show that using the TF- IDFVectorizer feature extraction is better than the CountVectorizer with the highest accuracy of 85.98% and for the final test the extraction feature with Cross Validation is better than not using Cross Validation with the highest accuracy value of 97.67%. Thus, testing the extraction feature that is best used is the TF-IDFVectorizer and by using the Cross Validation technique it can improve the performance of the Naïve Bayes model in the sentiment analysis of Indonesian-language twitter so that it.

Keywords : *Sentiment analysis, twitter, Naïve Bayes, feature extraction, Count Vectorizer, TF-IDF Vectorizer,Crosss Validation.*

I. Pendahuluan

Informasi merupakan sekumpulan sebuah data yang belum di kelolah untuk dijadikan informasi sebgai tujuan untuk kepentingan individu atau organisasi dalam mengambil keputusan. Media sosial digunakan untuk sarana komunikasi untuk individu dan kelompok, media sosial juga sering digunakan sebagai sarana informasiterbesar yang di gunakan oleh kebanyakan orang mengenai berita, berjualan, mencari seseorang dan lainnya. Ekstrasi

informasi merupakan teknik mengumpulkan informasi dari kumpulan data atau teks yang tidak terstruktur. Untuk memperoleh sebuah informasi dari data yang tidak terstruktur perlu didefinisikan terlebih dahulu sebagai informasi terstruktur yang akan diekstrak [1]. Dari data yang telah terkestrak didapatkan suatu informasi tentang opini atau pendapat dari pengguna media sosial terhadap entitas tertentu.

Analisa sentiment rumit di proses karena terdiri dari kata-kata gaul, ejaan yang salah, bentuk panjang pendeknya karakter, karakter yang berulang, penggunaan bahasa daerah dan penggunaan *emoticons*. Hal ini menyebabkan data opini harus diolah dengan dipilah dan mendata ulang secara keseluruhan dan detail. Prosesini dapat mengurangi efesiensi waktu, sehingga menimbulkan berkurangnya efesisensi dan efektivitas waktu kerja.

Beberapa penelitian data mining dan text mining yang telah dilakukan salah satunya dengan klasifikasi,teknik klasifikasi yang efektif dan sering diuji oleh para peneliti yaitu menggunakan metode *Naïve Bayes.* Klasifikasi teks atau opinion mining dalam analisa sentimen mempunyai masalah utama terhadap dimensi ruang klasifikasi cukup tinggi dari ruang fitur, biasanya terjadi pada teks yang memiliki puluhan ribu fitur serta data terlalu banyak noise pada data. Pemecahan masalah dari para peneliti dalam menanggapi ini dengan menggunakan fitur ekstrasi dan teknik validasi dapat mempengaruhi tingkat akurasi [5].

2. Tinjauan Studi

Perkembangan dunia teknologi informasi dan komunikasi yang pesat tidak terlepas dari penyedia layanan web yang menyediakan informasi yang beragam. Informasi yang menyebabkan pertambahan data yang kebanyakan berupa data teks dapat dijadikan sumber yang sangat potensial untuk digali lebih dalam. Salah satu contohnya adalah data text yang diambil dari twitter. Twitter adalah layanan jejaring sosial dan microblog daring yang memungkinkan penggunanya untuk mengirim dan membaca pesan berbasis teks hingga 140 karakter, yang dikenal dengan sebutan kicauan (tweet). Twitter didirikan pada bulan Maret 2006 oleh Jack Dorsey, dan situs jejaring sosialnya diluncurkan pada bulan Juli. Sejak diluncurkan, twitter telah menjadi salah satu dari sepuluh situs yang paling sering dikunjungi di internet, dan dijuluki dengan pesan singkat dari internet . Di twitter, penguna tak terdaftar hanya bisa membaca kicauan, sedangkan pengguna terdaftar bisa menulis kicauan melalui antarmuka situs web, pesan singkat (SMS) atau melalui berbagai aplikasi untuk perangkat seluler. Opini adalah pendapat, ide atau pikiran untuk menjelaskan kecenderungan atau preferensi tertentu terhadap perspektif dan ideologi akan tetapi bersifat tidak objektif karena belum mendapatkan pemastian atau pengujian, dapat pula merupakan sebuah pernyataan tentang sesuatu yang berlaku pada masa depan dan kebenaran atau kesalahannya serta tidak dapat langsung ditentukan misalnya menurut pembuktian melalui induksi.

Pada penelitian ini dataset yang digunakan berasal dari tweet atau komentar di twitter. Data yang diambil berdasarkan hastag berupa #baswedancoverbaswedan, isu ini muncul karena adanya pembelaan dari anies baswedan kepada novel baswedan menurut komentar di *twitter*. Data *tweet* diambil kemudian dilakukan pemrosesan menggunakan perangkat lunak *Python* tahap awal akan dilakukan *preprocessing* yang di dalam proses tersebut ada tahapan yang harus dilalui satu persatu, diawal proses dilakukan pemrosesan *tokenize* untuk menghilangkan nois, selanjutnya melakukan prose normalisasi kata-kata yang tidak baku, dan setelah itu melakukan penentuan label manual dengan paremeter positif = 1 dan negative = 2. Setelah dilakukan preprocess selanjutnya menggunakan feature extraction untuk mengekstrasi teks. Proses klasifikasi dengan menggunakan algoritma *naïve bayes. Objectives* pada penelitian ini adalah untuk mengetahui hasil klasifikasi berdasarkan model yang diusulkan dengan teknik *cross validation* dengan indikator yang ditentukan. Kemudian untuk tahap terakhir akan dilakukan pengukuran/ *measurement* evaluasi menggunakan *confusion matrix dan classification report* untuk melihat hasil akurasi algoritma.

Penelitian ini mengimplementasikan metode Naive Bayes Classifier kedalam suatu sistem untuk mengklasifikasikan data kedalam sentimen positif dan sentimen negatif berdasarkan data ulasan komentar film yang telah dikumpulkan. Sistem yang akan dibangun terdiri dari empat proses utama, yaitu pengumpulan data, preprocessing, pelabelan dan klasifikasi data. Data yang akan digunakan adalah data tweet komentar film berbahasa Indonesia. Data tweet ini diperoleh dengan membuat program scraping menggunakan library scrapy yang disediakan oleh python

2.1.1. Analisa Sentimen

Analisa sentimen atau opinion mining merupakan ilmu dibidang pengolahan bahasa secara alami untuk menganalisa data berbentuk opini dengan tujuan sebagai pendukung pengambilan keputusan [11].

Istilah analisis sentimen pertama kali didefinisikan pada tahun 2003 oleh Nasukawa dan Yi dengan penjelasan analisa sentiment merupakan penentu polaritas konektifitas (positif atau negatif) dan penentu kekuatan polaritas (sangat positif, sedikit positif , kurang positif) [12].

Tugas dasar analisa sentiment atau opinion mining adalah mengelompokkan kumpulan polaritas dari teks yang berada dalam dokumentasi, kalimat atau fitur entitas dengan tingkat aspek yang bersifat positif, negatif, atau netral [5].

2.1.2. Twitter

Twitter merupakan platform media sosial umum digunakan user untuk berkomunikasi dan menyebarkan informasi berupa tweets. *Tweets*

dapat di jadikan sumber data penting untuk melakukan penelitian *Neuro-Linguistic Programming* (NLP) seperti analisa sentimen, deteksi polaritas dan prediksi emoji[13].

Twitter menyediakan ribuan hingga jutaam data dari berbagai akun atau user di dunia, dengan data tersebut banyak kegiatan atau perkejaan dilakukan pada analisis data *twitter* seperti detiksi peristiwa *real-time*, analisa sentimen dan analisa hastag twitter [12].

Oleh karena itu *twitter* merupakan *platform* ideal untuk menunjukan berbeda dengan platform lain seperti Facebook, LinkedIn, dan MySpace, menurut *Wang Twitter* merupakan sebuah jejaring sosial yang dapatdigambarkan sebagai sebuah graph berarah, dengan demikian setiap pengguna bisa mengikuti penggguna lain tanpa harus izin dulu pada pemilik akunnya, dan si pengguna yang diikutin tidak harus mengikuti pengguna yang mengikutinya.

Para pengguna twitter bisa menerima dan mengirim *Tweets* melalui twitter ataupun beberapa applikasi yang kompatibel dengan twitter. Tanda hastag "#" biasa digunakan untuk menuliskan pesan sesuai dengan topik yang ada atau membuat topik baru, dan si pembalas bisa menggunakan @ untuk menyebut pengguna lain.

2.1.3. Python

Python sendiri merupakan bahasa Pemrograman yang sangat populer. Bahasa pemrograman lain dikenal sulit dan juga susah untuk dipahami, namun beda dengan Python. Python dikenal sebagai bahasa pemrograman yang lebih mudah untuk dipahami, sehingga Python sangat populer dan dapat dipelajari oleh banyak orang, mulai dari kaum awam ataupun yang sudah *expert* dan menguasai bahasa pemrograman. Python ini juga salah satu bahasa pemrograman yang paling disukai oleh *developer*, ilmuwan data, dan bahkan *hacker* karena fleksibilitas yang Python tawarkan. Google, YouTube, Dropbox adalah contohnya.

Python merupakan suatu perangkat lunak *open source* yang terkenal karena bahasa pemogramannya yang dinamis. *Python* untuk analitik disarankan menggunakan *Python* ver 3.0 ke atas dimana semua jenis API dapat digunakan saat melakukan klasfikasi atau analitik apapun dapat mudah digabungkan dan berfungsidengan baik [6].

2.1.4. Tweepy

Tweepy merupakan salah satu library dari python dalam mengakses API *twitter* untuk mengambil data dari *twitter*. Streaming API *tweepy* digunakan untuk mendapatkan *tweet* yang relevan sesuai kebutuhan [14]. *Crawling* menggunakan *tweepy* dapat menghemat waktu proses pengambilan data dan objek yang dapat bisa diatur sesuai kebutuhan peneliti.

2.1.5. Pandas

Pandas adalah pustaka open source berlisensi BSD yang menyediakan struktur data denganalat analisadata berkinerja tinggi dan mudah digunakan dalam bahasa pemograman *python*. Pandas diinisiasi pada tahun 2008 oleh *Wes McKinney* pada saat berada di lokasi *AQR Management Capital* [15].*Pandas* sering digunakan untuk membaca data atau memanggil data dari berbagai macam jenis data khususnya tekstual cara memanggil fungsi pandas kita bisa menuliskan "Import pandas as pd".

2.1.6. Regular Expression (Regex)

Regex adalah urutan dalam pendefinisian pola pencarian karakter, dalam artian lain regex digunakan dalam pengenalan karakter untuk mendapatkan data yang bersih dari data ambigu. *Regex* dapat digunakan dengan modul "re" atau "import re" pada *python* [16].

2.1.7. TFScikit-Learn (Sklearn)

Sklearn adalah salah satu modul python yang mengintegrasikan banyak algoritma dalam pembelajaran mesin. Pustaka sklearn awalnya dikembangkan oleh Cournapeu pada tahun 2007, akan tetapi untuk rilisan pertama secara resmi pada tahun 201 Sklearn merupakan bagian dari pustaka *SciPy (Science Python)*, dimana satu set pustaka ini dibuat untuk komputasi ilmiah khususnya analisa data [17].

2.1.8. Count-Vectorizer (CV)

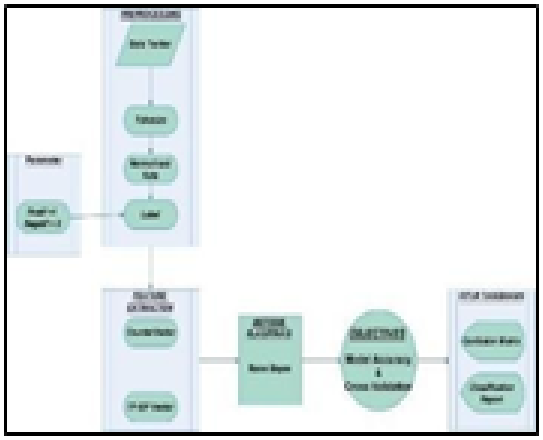
Count–Vectorizer merupakan salah satu teknik yang berdasarkan kemunculan kata dalam dokumen. Dalam teknik ini bisa perhitungan tokenisasi yang dilakukan serta banyak parameter lain yang dapat menyumparnakan jenis fitur salah satunya seperti generate Ngram. CV juga menghitung jumlah kata yang sering muncul tetapi sering sekali jumlah kata yang jarang muncul akan tertupi walaupun kata tersebut bisa menjadi hal penting dalam fitur dokumen [18]. Batasan CV ini dapat ditangani dengan menggunakan teknik TF-IDF.

2.1.9. Data Mining

Data mining atau penambangan data adalah proses penyaringan data yang sesuai dari kumpulan data yang banyak atau besar dengan menggunakan teknik dan algoritma yang berbeda seperti asosiasi, pengelompokkan, dan klasifikasi yang tujuan untuk kepentingan pengambilan keputusan. Penambangan data pengetahuan dari data berjumlah besar juga didefinisikan sebagai menemukan informasi tersembunyi dari *database* [22].

3. Desain Penelitian/Metodologi

Dalam pelaksanaan penelitian ini, terdapat beberapa langkah sampai akhirnya menerapkan metode klasifikasi *Naïve Bayes* dan pengujian metode. Secara umum langkah-langkah tersebut digambarkan pada diagram alur dibawah ini:



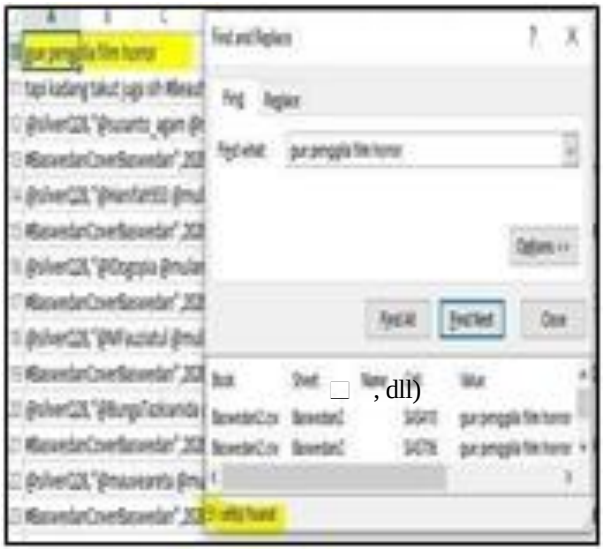
Gambar 1. Metode Penelitian

3.1. Preprocessing

Dalam membuat keputusan yang baik, maka harus menggunakan data yang baik pula (lengkap, benar, konsisten, dan terintegrasi). Sebelum melakukan data mining, perlu dilakukan *preprocessing* data terlebih dahulu untuk memastikan data yang akan diolah di *data mining* adalah data yang baik. Data yang kualitasnya kurang baik, dapat disebabkan oleh beberapa hal, yaitu: Data tidak lengkap, noisy (ada data yang berbeda sendiri), tidak konsisten (tidak sesuai dengan rule yang ditentukan). Untuk mengatasi masalah tersebut, maka dilakukanlah *preprocessing* data sebelum diolah dengan data mining**,** proses ini dilakukan menggunakan *script python* dan secara manual pada tahapan ini terdiri dari beberapa tahap sebagai berikut:

3.1.1 Remove Duplicate

Pada proses ini bertujuan untuk menyeleksi komentar *tweet* yang berulang. Sebab pada *twitter* terdapat fitur *retweet* yang memberikan dampak banyaknya teks berulang dengan topik dan isi yang sama. Hal ini menyebabkan banyaknya dimensi yang harus diproses saat pembentukan model dan mengakibatkan lamanya waktu pemrosesandata, contoh data duplikasi sebagai berikut.



Gambar 2. Data Duplikasi

3.1.2 Case Folding

Pada proses *case folding* ini, semua data tweet atau komentar akan dirubah menjadi lower case atau huruf kecil dengan menggunakan fungsi *lower* () dari *library regex* (*import re*). Berikut contoh hasil dari tahap*case folding* :

Tabel 1. Contoh Proses *Lower*

Input	Output
"Melihat kisruh dinegara ini seolah tiada akhir, gw	“kisruh dinegara ini seolah tiada akhir, gw pengen bgt rasanya pindah kewarganegaraan.

Data Mining memiliki suatu rangkaian proses yang harus dilakukan sebelum dapat memperoleh informasi baru. Tahap-tahap dalam data mining adalah sebagai berikut [4]:

a. Data cleaning
   Pembersihan dari merupakan proses menghilangkan *noise* dan data yang tidak konsisten.
b. Data integration
   Proses dimana menggabungkan data dari berbagai macam sumber data. Proses ini dilakukan ketika menggunakan sumber data yang lebih dari satu.
c. Data selection
   Proses menyeleksi data dimana data yang akan digunakan dalam proses *data mining* diambil dan membiarkan data yang tidak digunakan.
d. Data transformation
   Proses mengubah data ke dalam bentuk yang dapat digunakan dalam perhitungan suatu algoritma
e. Data mining
   Proses menemukan pola dari dataset yang digunakan sebagai basis pengetahuan.
f. Pattern evaluation
   Merupakan proses menganalisis hasil dari proses mining menggunakan suatu satuan ukur.
g. Knowledge presentation
   Merupakan proses untuk menampilkan hasil dari proses *mining*.

3.2 Cleansing atau Tokenizing

Pada tahap cleansing ini akan dilakukan proses penghapusan kata, karakter dan simbol yang tidakdiperlukan untuk mendapatkan data yang lebih bersih, contoh kata, karakter dan simbol sebagai berikut:

1. Karakter HTML (<, >, dll)
2. Ikon emosi (:0,
3. Hastag (#)
4. Usernama (@username)
5. Url (*http://website.com*)
6. Email (nama@website.com)
7. Tanda baca atau *punctuation* (“,”,”.”,”?”,dll)

8. Retweet (RT)
9. Angka.

4. Pembahasan

Pada penelitian ini jumlah dataset yang digunakan sebanyak 1069 data *tweet* yang terdiri tweet positif, netral dan negatif dari pengambilan *crawling* data dengan hastag #baswedancoverbaswedan sebagai objek utama penelitiannya. Data tweet diperoleh pada tanggal 18 Juni 202 Proses pengumpulan data dilakukan dengan proses *crawling* data yang menggunakan perangkat lunak *python* dengan *library tweepy*. Jumlah data yang diperoleh dari proses tersebut sebanyak 5406 data. Setelah di lakukan proses preprocessing atau seleksi data, dataset yang diperoleh sebanyak 1069 data.

Pengujian dilakukan dengan membagi dataset dengan ratio 80%:20%. Proses akan diuji menggunakan metode *Naïve Bayes* yaitu model *Multinomial Naïve Bayes, Gaussian Naïve Bayes, Bernoulli Naive Bayes, Naïve Bayes* dengan tfidf vector dan terakhir proses pengujian *Naïve Bayes* menggunakan teknik *cross validation*. Hasil pengujian akan ditunjukan dengan akurasi, *confusion matrix*, recall, dan presisi.

Akurasi yang didapatkan menghasilkan seberapa dekat nilai prediksi dengan nilai sebenarnya. Berdasarkan pengujian di atas dapat dilihat nilai performa atau akurasi dari pengujian menggunakan metode *Multinominal Naïve Bayes Tfidf* mendapatkan akurasi tertinggi sebesar 85.98% sedangkan dengan menggunakan metode *Multinominal Naïve Bayes Tfidf Cross Validation* mendapatkan akurasi tertinggi sebesar 97.67%. Akurasi yang didapatkan menunjukan selisih 11.69% . Hal ini menunjukan bahwa dengan menggunakan teknik *cross validation* dapat meningkatkan akurasi dari metode *Naïve Bayes* walaupun tidak terlalu signifikan.

5. Penutup

Berdasarkan penelitian yang telah di lakukan mendapatkan hasil dari pengujian sebagai berikut:

5.1 Pengujian menggunakan fitur ekstrasi teks *TF- IDFVectorizer* terhadap model *Naïve Bayes* (*Multinominal Naïve Bayes*) menunjukan hasil yang lebih baik dari pengujian *CountVectorizer* dengan ditunjukan akurasi sebesar 85.98%.

5.2 Proses pengujian akhir yaitu menggunakan validasi dengan teknik *Cross Validation* menunjukan hasil yang lebih baik dibandingkan pengujian yang tidak menggunakan teknik *Cross Validation*, nilai akurasi tertinggi yang di dapat sebesar 97.67%.

Dengan ini dapat disimpulkan bahwa pengujian fitur ekstrasi yang baik digunakan adalah Vectorizer dan dengan teknik *Cross Validation* dapat membantu dalam meningkatkan kinerja model *Naïve Bayes* pada analisis sentimen tweet berbahasa Indonesia dengan mendapatkan selisih nilai akurasi sebesar 11.69%.

Daftar Pustaka

[1] Zy, A. T., & Nugroho, A. 2018. Comparison Algorithm Classification *Naive Bayes*, Decision Tree, and Neural Network for Analysis Sentiment. International Conference on Economic, Business, and Accounting, 1(c), 115–115.

[2] G. Gupta and G. S. Bhathal, SENTIMENT ANALYSIS OF ENGLISH TWEETS USING DATA MINING: Data Mining, Sentiment Analysis. BookRix, 2018

[3] H. Sumarno, "Komparasi algoritma klasifikasi machine learning pada analisis sentimen film berbahasa Indonesia," vol. 4, no. 2, pp. 189–196, 2017.

[4] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," INTEGER J. Inf. Technol., vol. 1, no. 1, pp. 32–41, 2017, [Online].

[5] K. J. Prayoga, A. Nugroho, and N. Wiyatno, "Komparasi feature selection Particle Swarm Optimization (pso) dengan Genetic Algorithm (ga) terhadap algoritma Naïve Bayes pada analisis sentimen twitter," Pros. Semin. Nas. Teknol. dan Sains,no. September, 2019.

[6] A. Goel, J. Gautam, and S. Kumar, "Real time sentiment analysis of tweets using Naive Bayes," Proc. 2016 2nd Int. Conf. Next Gener. Comput. Technol. NGCT 2016, no. October, pp. 257–261, 2017, doi: 11109/NGCT.2016.7877424.

[7] A. G. Prasad, S. Sanjana, S. M. Bhat, and B. S. Harish, "Sentiment analysis for sarcasm detection on streaming short text data," *2017 2nd Int. Conf. Knowl. Eng. Appl. ICKEA 2017*, vol. 2017- January, no. 2009, pp. 1–5, 2017, doi: 11109/ICKEA.2017.8169892.

[8] S. Elbagir and J. Yang, "Sentiment analysis of twitter data using machine learning techniques and scikit-learn," ACM Int. Conf. Proceeding Ser., 2018, doi: 11145/3302425.3302492.

[9] E. Miranda, M. Aryuni, R. Hariyanto, and E. S. Surya, "Sentiment Analysis using Sentiwordnet and Machine Learning Approach (Indonesia general election opinion from the twitter content)," Proc. 2019 Int. Conf. Inf. Manag. Technol. ICIMTech 2019, vol. 1, no. August, pp. 62–67, 2019, doi: 11109/ICIMTech.2019.8843734.

[10] N. A. Lestari and J. Timur, "METODE NAÏVE BAYES CLASSIFIER DENGAN TEXTBLOB UNTUKANALISIS SENTIMEN TERHADAP PELAYANAN," vol. 4, no.
September, 202

[11] D. K. Tayal and S. K. Yadav, "Sentiment analysis on social campaign 'Swachh Bharat Abhiyan' using unigram method," AI Soc., vol. 32, no. 4, pp. 633– 645, 2017, doi: 11007/s00146- 016-0672-5.

[12] A. Yadav, C. K. Jha, A. Sharan, and V. Vaish, "ScienceDirect ScienceDirect Sentiment analysis of financial news using unsupervised approach Sentiment analysis of financial news using unsupervised approach,"Procedia Comput. Sci., vol. 167, no. 2019, pp. 589–598, 2020, doi: 11016/j.procs.20203.325.

[13] N. Choudhary, R. Singh, V. Anvesh Rao, and M. Shrivastava, "Twitter corpus of Resource-Scarce Languages for Sentiment Analysis and Multilingual Emoji Prediction," Proc. 27th Int. Conf. Comput. Linguist., pp. 1570– 1577, 2018.

[14] S. Mundra, A. Dhingra, A. Kapur, and D. Joshi, Prediction of a movie's successusing data mining techniques, vol. 106. Springer Singapore, 2019.

[15] U. Ziegenhagen, "Using Python for Research," 2016, [Online]. Available: https://www.edx.org/course/using-python-research-harvardx- ph526x.

[16] "FUNDAMENTAL SCIENCES AND APPLICATIONS," vol. 24, 2018.

[17] F. Nelli, *Python data analytics: With Pandas, NumPy, and Matplotlib: Second edition.* 2018.