

Perbandingan Metode *Naïve Bayes* dan *Support Vector Machine* pada Analisis Sentimen *Twitter*

Mujaddid Izzul Fikri¹, Trifebi Shina Sabrila², Yufis Azhar³

^{1, 2, 3}Universitas Muhammadiyah Malang

¹mujaddidizzulfikrik.mif@gmail.com, ²trifebiss@gmail.com, ³yufis @umm.ac.id

ABSTRAK

Twitter merupakan salah satu media sosial yang banyak digunakan oleh masyarakat sebagai media komunikasi dan memperoleh informasi. Melalui media sosial ini, pengguna dapat menyampaikan berbagai macam opini maupun komentar terhadap suatu isu. Opini dan komentar yang pengguna sampaikan melalui *tweets* yang dituliskannya pun dapat digunakan untuk analisis sentimen. Maka dari itu, dalam penelitian ini dilakukan analisis sentimen terhadap *tweets* yang berhubungan dengan Universitas Muhammadiyah Malang (UMM) untuk mengetahui opini masyarakat mengenai kampus ini. Analisis dilakukan dengan mengklasifikasikan *tweets* yang berisi sentimen masyarakat mengenai UMM. Metode klasifikasi yang digunakan dalam penelitian ini adalah *Naïve Bayes* dan *Support Vector Machine* (SVM) dengan pembobotan menggunakan TF-IDF. Hasil komparasi kedua metode menunjukkan bahwa *Naïve Bayes* mendapatkan hasil akurasi yang lebih baik dari SVM dengan akurasi sebesar 73,65%.

Kata Kunci: Analisis Sentimen, *Naïve Bayes*, *Support Vector Machine*

ABSTRACT

Twitter is one of the social media that is widely used by the public as a communication media and obtain information. Through this social media, users can submit various opinions or comments on an issue. The opinions and comments that users submit through the tweets they send can be used for sentiment analysis. Therefore, in this study sentiment analysis of tweets related to the University of Muhammadiyah Malang (UMM) was carried out to determine public opinion about this campus. The analysis was carried out by classifying tweets that contain people's sentiments regarding UMM. The classification method used in this study is Naïve Bayes and Support Vector Machine (SVM) by weighting the term using TF-IDF. The result of the two methods shows that Naïve Bayes gets better accuracy than SVM with an accuracy of 73,65%

Keywords: Sentiment Analysis, *Naïve Bayes*, *Support Vector Machine*

1. PENDAHULUAN

Twitter merupakan salah satu media sosial terpopuler yang berperan sebagai wadah komunikasi di masyarakat. Dengan menggunakan *Twitter*, seluruh orang di dunia dapat terhubung dengan keluarga, teman, dan kerabat melalui komputer atau ponsel mereka. Salah satu layanan yang disediakan oleh *Twitter* kepada penggunaannya adalah pembuatan pesan status (disebut "*tweets*") yang dapat dibaca oleh pengguna *Twitter* lainnya dan biasanya berisi ungkapan pendapat pengguna dalam berbagai topik dengan batasan sebanyak 140 karakter, sehingga *twitter* menjadi salah satu situs yang menyediakan kumpulan data opini dari masyarakat di seluruh dunia [1], [2], [3].

Analisis sentimen termasuk ke dalam salah satu bidang dari *Natural Language Processing* (NLP) dan merupakan suatu proses yang digunakan untuk membantu mengidentifikasi isi dari dataset yang berupa opini atau pandangan (sentimen) berbentuk teks terhadap suatu isu atau kejadian bersifat positif, negatif atau netral [4], [5]. Analisis sentimen merupakan bidang penelitian yang lumayan populer dan dianggap mampu memberikan keuntungan dalam berbagai aspek. Umumnya,

analisis sentimen diterapkan dalam berbagai aspek, seperti prediksi harga saham, isu politik, kepuasan terhadap suatu produk atau layanan, analisis reputasi, dan sebagainya [5]. Salah satu analisis sentimen yang dapat dilakukan adalah analisis sentimen terhadap suatu isu yang terdapat pada *Twitter*.

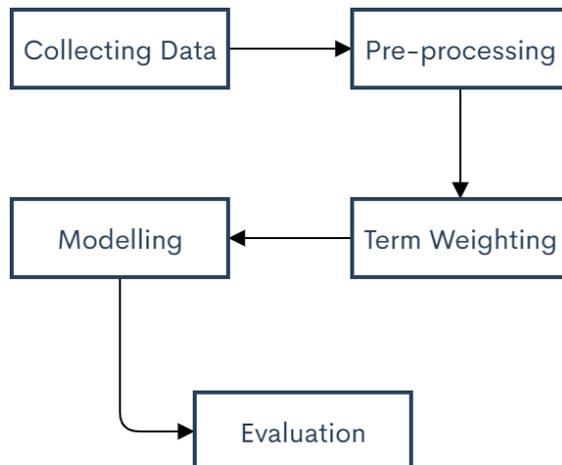
Tweets yang merupakan hasil dari penyaluran opini dan komentar, merupakan *resource* yang dapat digunakan untuk menganalisis sentimen khalayak umum terhadap suatu instansi maupun perorangan. Hal tersebut disebabkan karena *tweets* tersebut mengandung sentimen yang dapat dijadikan sebagai tolak ukur pandangan khalayak umum yang dapat dijadikan sebagai bahan evaluasi ke depannya. Dalam menentukan sentimen suatu *tweet*, dapat dilakukan dengan mengklasifikasikannya ke dalam tiga kelas, yaitu positif, negatif, dan netral [5]. Dalam penelitian kali ini, analisis sentimen dilakukan terhadap Universitas Muhammadiyah Malang (UMM). Oleh karena itu, penelitian ini bertujuan untuk mengetahui opini masyarakat mengenai kampus tersebut dengan menganalisa sentimen *tweet* yang berkaitan dengan Universitas Muhammadiyah

Malang dengan klasifikasi menggunakan metode *Naïve Bayes* dan *Support Vector Machine (SVM)*.

Pada penelitian sebelumnya yang dilakukan oleh Alec Go, dll, analisis sentimen dilakukan dengan menggunakan satu jenis algoritma klasifikasi yaitu *naive bayes*. Namun, hasil sentimen hanya diklasifikasikan ke dalam kategori positif dan negatif [2]. Sedangkan *tweets* tidak hanya terbagi dalam dua kategori tersebut. Saat menuliskan sebuah *tweet*, seseorang tidak hanya dalam kondisi perasaan yang baik, ataupun buruk. Penelitian lain juga dilakukan oleh Winda Estu Nurjanah, dkk, pada penelitiannya dilakukan analisis sentimen dengan menggunakan metode *K-Nearest Neighbor* tanpa dilakukan perbandingan dengan metode lainnya, pembobotan pun dilakukan dengan berdasarkan jumlah *retweet* [6]. Oleh karena itu, dalam penelitian ini terdapat satu kategori sentimen dalam klasifikasi sentimennya, yaitu netral. Dengan adanya tiga kategori sentimen, diharapkan dapat memberikan hasil yang lebih baik dalam merepresentasikan hasil sentimen dari datasetnya. Dan untuk hasilnya, akan dilakukan perbandingan antara dua metode klasifikasi yaitu *Naive Bayes* dan *SVM*.

2. METODOLOGI PENELITIAN

Pada bagian ini akan dijelaskan mengenai perancangan sistem dalam penelitian ini. Metode dalam penelitian ini dilakukan melalui beberapa tahap. Secara garis besar, alur penelitian dapat dijelaskan melalui diagram pada Gambar 1.



Gambar 1. Alur Desain Sistem

Collecting Data

Data yang digunakan dalam penelitian ini adalah *tweets* yang mengandung kata “Universitas Muhammadiyah Malang”, “muhammadiyah”, “UMM”, atau “unmu” selama tahun 2018 hingga 2019. Proses pengumpulan data dilakukan dengan teknik *scraping* menggunakan *twitterscraper* yang

bisa langsung diinstal dan digunakan melalui terminal atau cmd. Dari hasil *scraping*, didapatkan sebanyak 2654 data *tweets* yang disimpan ke dalam format .csv.

Pre-processing

Pre-processing adalah proses mengubah data mentah menjadi data yang sesuai dengan prosedur *mining* yang akan dilakukan dan merupakan tahap yang paling penting dalam *data mining* [7]. Pada tahap ini dilakukan proses pembersihan pada data yang masih kotor. Proses pembersihan yang dilakukan pada penelitian ini antara lain:

Special Character Removal

Tweets yang didapatkan dari twitter mengandung berbagai macam tipe informasi serta berbagai macam karakter spesial yang tidak diperlukan dalam proses analisis sentimen. Jika ditemukan data yang mengandung suatu link, maka akan dilakukan proses penghapusan URL. Selain itu, proses penghapusan juga dilakukan terhadap karakter spesial lainnya seperti hastag ataupun username dari suatu user (dituliskan @username).

Case Folding

Case folding merupakan proses yang dilakukan untuk mengubah setiap kata yang ada di dalam dataset menjadi huruf kecil menggunakan fungsi *lowercase*.

Stopword Removal

Stopword removal merupakan proses penghapusan kata-kata yang dianggap kurang/tidak memiliki makna untuk mempengaruhi sentimen. Kata-kata yang dimaksud seperti “yang”, “di”, “dan”, “pada”, dan sebagainya.

Tokenization

Tokenization adalah proses pemecahan teks dari sebuah kalimat menjadi potongan-potongan kata.

Term Weighting

Semua fitur yang telah diekstraksi, kemudian ditransformasikan ke *Bag-of-Words (BoW)* untuk menentukan bobot pada setiap katanya yang disebut dengan *term weighting*. *BoW* pada tahapan ini, merupakan matriks yang memetakan setiap fitur kata terhadap bobotnya. Sehingga menghasilkan matrix yang memiliki baris sejumlah *tweet*nya, dan memiliki kolom sebanyak jumlah kata unik dari keseluruhan *tweet*. Pada penelitian kali ini pembobotan kata dihitung menggunakan teknik *TF-IDF (Term Frequency-Inverse Document Frequency)*.

TF-IDF

TF-IDF (*Term Frequency-Inverse Document Frequency*) merupakan suatu metode algoritma yang memberikan bobot terhadap teks. TF adalah frekuensi suatu kata muncul dalam dokumen, sedangkan IDF adalah nilai *invers* dari dokumen yang mengandung kata tersebut. TF dan IDF akan dikalikan sehingga menghasilkan nilai bobot dari kata tersebut [8]. Berikut persamaan TF-IDF:

$$TF - IDF(d, t) = TD(d, t) * IDF(t) \quad (1)$$

dimana:

$$TF(d, t) = \frac{\text{Jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d}$$

$$IDF(t) = \log \frac{\text{total dokumen}}{\text{jumlah dokumen yang mengandung kata } t}$$

dan:

t = kata

d = dokumen

Modelling

Setelah didapatkan bobot pada setiap *feature* (kata), selanjutnya akan dilakukan proses pemodelan *data mining*. *Data mining* merupakan kegiatan yang berfokus pada bagaimana mengekstrak informasi yang penting dari suatu dataset dan menyajikannya ke dalam bentuk yang mudah dipahami oleh *user*. Terdapat dua jenis *data mining*, yaitu *predictive data mining* dan *descriptive data mining*. Metode *data mining* yang akan digunakan dalam penelitian ini adalah metode klasifikasi. Klasifikasi sendiri merupakan salah satu contoh dari *predictive data mining*. Pada *predictive data mining*, terdapat proses *data splitting*. *Data splitting* merupakan proses membagi data menjadi *data training* (data latih) untuk mengajari sistem dan *data testing* (data uji) untuk menguji sistem. Algoritma *Naive Bayes* dan SVM yang digunakan dalam penelitian ini, termasuk ke dalam *supervised learning*. Sehingga tiap data yang digunakan untuk *Training* model adalah yang telah diberikan label. Dalam penelitian ini, label tersebut adalah tiga kelas sentimennya.

Naive Bayes

Algoritma *Naive Bayes* merupakan salah satu algoritma yang terdapat dalam teknik klasifikasi. *Naive Bayes* merupakan pengklasifikasian dengan bentuk model probabilistik dan statistik yang disederhanakan dengan berdasar pada teorema Bayes dengan asumsi bahwa setiap atribut bersifat bebas (*independence*). Dengan kata lain, algoritma ini mengasumsikan bahwa ada atau tidak ciri tertentu

dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya [9], [10], [11].

Berikut persamaan dari teorema Bayes [11]:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (2)$$

dimana:

X : Data dengan class yang belum diketahui

C : Hipotesis data X merupakan suatu kelas spesifik
 $P(C|X)$: Probabilitas hipotesis C berdasarkan kondisi X (*posterior probability*)

$P(C)$: Probabilitas hipotesis C (*prior probability*)

$P(X|C)$: Probabilitas X berdasarkan kondisi pada hipotesis C (*likelihood*)

$P(X)$: Probabilitas X (*predictor prior probability*)

Support Vector Machine (SVM)

SVM adalah algoritma *Machine Learning* yang menerapkan fungsi *hyperplane* pada data sehingga terbentuk daerah-daerah tiap kelas. *Hyperplane* sendiri merupakan sebuah fungsi yang digunakan sebagai pemisah antar kelas yang ada. Dalam memprediksi suatu kelas dari data, SVM akan melabelinya berdasarkan daerah kelas mana yang merupakan tempat dari data tersebut. SVM biasanya digunakan pada dataset besar yang diambil dari situs *online* dan menjadi populer karena penerapannya dalam klasifikasi teks [12]. Prinsip SVM adalah membangun *hyperplane* yang memiliki ukuran margin yang sama dan tidak cenderung mendekati daerah dari salah satu kelas. Hal tersebut dapat dilakukan dengan pengukuran *margin* kemudian dilakukan pencarian titik maksimalnya. Usaha pada pencarian *hyperplane* terbaik sebagai pemisah antar kelas merupakan inti dari metode SVM.

Metode SVM sendiri memiliki 3 model pendekatan atau yang lebih sering disebut dengan kernel, yaitu kernel radial, linier, dan sigmoid. Pada penelitian ini, model pendekatan yang digunakan adalah kernel linier.

Evaluation

Evaluasi dalam penelitian ini bertujuan untuk mengetahui kinerja atau performa dari model yang diusulkan. Metode yang digunakan untuk evaluasi pada penelitian ini adalah *confusion matrix*. *Confusion matrix* merupakan sebuah tabel yang memberikan informasi perbandingan hasil klasifikasi yang dilakukan oleh sistem (prediksi) dengan hasil klasifikasi yang sebenarnya. Tabel pada *confusion matrix* menunjukkan jumlah data uji yang diklasifikasikan dengan benar dan jumlah data uji yang salah diklasifikasikan. Contoh *confusion matrix 2x2* dan *confusion matrix 3x3* dapat dilihat pada Gambar 2 dan Gambar 3.

		Actual Class	
		Yes	No
Predicted Class	Yes	TP	FN
	No	FP	TN

Gambar 2. Contoh *confusion matrix* 2x2

		Predicted Class		
		A	B	C
Actual Class	A	AA	AB	AC
	B	BA	BB	BC
	C	CA	CB	CC

Gambar 3. Contoh *confusion matrix* 3x3

Dalam *confusion matrix*, terdapat 4 istilah nilai hasil klasifikasi yang nantinya akan digunakan untuk menghitung nilai akurasi, presisi, *recall*, dan *f1-score*, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN).

Nilai *True Positive* (TP) dan *True Negative* (TN) adalah hasil klasifikasi yang benar. Nilai *False Positive* (FP) merupakan nilai dimana hasilnya diprediksi *positive* namun sebenarnya *negative*, sedangkan *False Negative* (FN) adalah nilai dimana hasil prediksinya *negative* namun sebenarnya *positive* [1].

Berdasarkan nilai dari *confusion matrix* tersebut, dapat dihasilkan nilai akurasi, presisi, *recall*, dan *f1-score* dengan persamaan berikut [1]:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (3)$$

$$Presisi = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$f1 - score = \frac{2 \times Presisi \times Recall}{Presisi + Recall} \quad (6)$$

Persamaan di atas dapat digunakan pada contoh *confusion matrix* 2x2, sedangkan untuk *confusion matrix* 3x3, persamaan yang dapat digunakan untuk menghasilkan nilai akurasi, presisi, *recall* dan *f1-score* adalah sebagai berikut [13]:

$$Akurasi = \frac{AA+BB+CC}{AA+AB+AC+BA+BB+BC+CA+CB+CC} \quad (7)$$

$$Presisi = \frac{\left(\frac{AA}{AA+BA+CA} + \frac{BB}{AB+BB+CB} + \frac{CC}{AB+BB+CB}\right)}{3} \quad (8)$$

$$Recall = \frac{\left(\frac{AA}{AA+AB+AC} + \frac{BB}{BA+BB+BC} + \frac{CC}{CA+CB+CC}\right)}{3} \quad (9)$$

$$f1 - score = \frac{2 \times Presisi \times Recall}{Presisi + Recall} \quad (10)$$

3. HASIL DAN PEMBAHASAN

Dataset

Dataset yang digunakan dalam penelitian ini adalah *dataset* yang bersumber dari Twitter berupa *tweets* yang mengandung kata “UMM”, “Universitas Muhammadiyah Malang” ataupun “unmuh” pada tahun 2018 dan 2019. Pengumpulan data dilakukan dengan teknik *scraping* menggunakan *twitterscraper*.

Dataset yang dikumpulkan berjumlah 2654 *record*. Namun, setelah dilakukan pembersihan tersisa sebanyak 2030 *record*. Kemudian dilakukan proses labelisasi oleh enam anotator yang terdiri dari lima anotator primer yang menentukan termasuk sentimen apakah suatu *tweet*, dan satu anotator sekunder sebagai pelaksana *quality control* yang memverifikasi apakah hasil dari pelabelan yang dilakukan oleh anotator primer sudah tepat atau belum.

Dari proses pelabelan, didapatkan hasil bahwa *tweet* yang termasuk ke dalam *class* netral sebanyak 749 *tweet*, *class* positif sebanyak 702 *tweet*, dan *class* negatif sebanyak 579 *tweet*.

Dari 2030 *record* yang digunakan, dilakukan *data split* untuk membagi data ke dalam dua model, yaitu *data training* (data latih) dan *data testing* (data uji). *Data training* berjumlah sebanyak 1624 data dan *data testing* berjumlah sebanyak 406 data.

Hasil

Pada bagian ini akan dijelaskan hasil pengujian terhadap model usulan yang telah dibuat. Pengujian dilakukan sesuai dengan tahap yang sudah dijelaskan pada bagian metodologi penelitian. Algoritma yang diusulkan dalam model diimplementasikan dengan menggunakan Bahasa pemrograman *python* dan *machine learning library* “sklearn”. Hasil dari pengujian kemudian akan dievaluasi dengan pengukuran akurasi dan presisi sistem.

Tabel perbandingan antara kelas yang sebenarnya (*actual label*) dengan hasil prediksi (*prediction*) metode *Naïve Bayes* dan *SVM* terhadap *data testing* dapat dilihat pada Tabel 1.

Tabel 1. Perbandingan jumlah *actual label* dengan *prediction*

Label		Naive Bayes	SVM
Netral	Actual	149	149
	Prediction	159	181
Positif	Actual	141	141
	Prediction	126	116
Negatif	Actual	116	116
	Prediction	121	109

Dari hasil pengujian pada Tabel 1, menunjukkan bahwa model yang menggunakan metode *Naive Bayes*, memprediksi dengan benar lebih banyak sentimen dibandingkan metode *SVM*. Metode *Naive Bayes* memprediksi sebanyak 299 sentimen dari *tweet* dengan benar, lebih banyak 14 prediksi benar dibandingkan dengan metode *SVM* yang memprediksi 285.

Untuk hasil yang lebih lengkapnya, dapat dilihat pada *Confusion matrix* dari pengujian sistem dengan menggunakan algoritma *Naive Bayes* dan *SVM* yang terdapat dalam Gambar 4 dan Gambar 5.

		Predicted Class		
		pred. neutral	pred. positive	pred. negative
Actual Class	true neutral	103	39	7
	true positive	51	86	4
	true negative	5	1	110

Gambar 4. *Confusion matrix* dengan algoritma *Naive Bayes*

		Predicted Class		
		pred. neutral	pred. positive	pred. negative
Actual Class	true neutral	106	40	3
	true positive	66	74	1
	true negative	9	2	105

Gambar 5. *Confusion matrix* dengan algoritma *SVM*

Selanjutnya, berdasarkan nilai pada *confusion matrix* pada Gambar 4 dan Gambar 5, didapatkan perbandingan hasil perhitungan akurasi dan presisi sistem yang diperlihatkan pada Tabel 2.

Tabel 2. Perbandingan hasil perhitungan akurasi, presisi, *recall*, dan *f1-score*

Evaluasi	Naive Bayes	SVM
Akurasi	73,65%	70,20%
Presisi	0,75	0,73
<i>Recall</i>	0,75	0,71
<i>f1-score</i>	0,75	0,72

Hasil pengujian tersebut menunjukkan bahwa metode *Naive Bayes* memiliki hasil akurasi, presisi, *recall*, dan *f1-score* yang lebih unggul sebesar 3,45%, 0,02, 0,04, dan 0,03 dibandingkan dengan metode *SVM*.

Pembahasan

Sentimen positif yang terkait dengan UMM merupakan *tweet-tweet* yang pada umumnya mengapresiasi program kerja kampus dan prestasi dari mahasiswanya, sedangkan pada sentimen negatif, berisi tuntutan atau keluhan mahasiswa seperti sistem administrasi yang bermasalah atau adanya infrastruktur kampus yang kurang maksimal. Dan untuk sentimen netral umumnya berisi *update-an tweet-tweet* kegiatan yang dilakukan dan berkaitan dengan UMM. Sentimen yang positif, dapat dijadikan sebagai tolak ukur kepuasan massa terhadap prestasi-prestasi UMM sedangkan untuk sentimen yang negatif dapat dijadikan sebagai sarana penampungan kritik dan saran untuk melakukan tindakan yang dapat meminimalisir kesalahan ke depannya.

4. KESIMPULAN DAN SARAN

Dari hasil klasifikasi analisis sentimen terhadap dataset *tweets* yang berkaitan dengan Universitas Muhammadiyah Malang, dapat disimpulkan bahwa kedua metode dapat melakukan performa yang baik dan metode *Naive Bayes* memiliki hasil akurasi, presisi, *recall*, dan *F1-Score* yang lebih baik dibandingkan dengan metode *SVM*.

Dalam penelitian ini menggunakan metode *Naive Bayes* dan *SVM* untuk proses klasifikasi. Selain kedua metode tersebut, dapat juga dicoba berbagai metode klasifikasi yang lain seperti *K-Nearest Neighbor*, *Decision Tree*, *rule-based*, dll untuk analisis sentimen.

5. REFERENSI

- [1] A. Fathan Hidayatullah and A. Sn, "ISSN: 1979-2328 UPN "Veteran," *Semin. Nas. Inform.*, vol. 2014, no. semnasIF, pp. 115–122, 2014.
- [2] A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification using Distant Supervision," *Processing*, vol., pp. 1–6, 2009.
- [3] H. Parveen and S. Pandey, "Sentiment analysis on Twitter Data-set using Naive Bayes algorithm," *Proc. 2016 2nd Int. Conf. Appl. Theor. Comput. Commun. Technol. iCATccT 2016*, pp. 416–419, 2017, doi: 10.1109/ICATCCT.2016.7912034.
- [4] S. Fanissa, M. A. Fauzi, and S. Adinugroho, "Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking | Jurnal Pengembangan Teknologi Informasi

- dan Ilmu Komputer,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 8, pp. 2766–2770, 2018.
- [5] V. Chandani, F. I. Komputer, and U. D. Nuswantoro, “Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film,” *J. Intell. Syst.*, vol. 1, no. 1, pp. 56–60, 2015.
- [6] W. E. Nurjanah, R. S. Perdana, and M. A. Fauzi, “Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 1, no. 12, pp. 1750–1757, 2017, doi: 10.1074/jbc.M209498200.
- [7] S. A. Alasadi and W. S. Bhaya, “Review of data preprocessing techniques in data mining,” *J. Eng. Appl. Sci.*, vol. 12, no. 16, pp. 4102–4107, 2017, doi: 10.3923/jeasci.2017.4102.4107.
- [8] B. K. Hananto, A. Pinandito, and A. P. Kharisma, “Penerapan Maximum TF-IDF Normalization Terhadap Metode KNN Untuk Klasifikasi Dataset Multiclass Panichella Pada Review Aplikasi Mobile,” vol. 2, no. 12, pp. 6812–6823, 2018.
- [9] B. Olabenjo, “Applying Naïve Bayes Classification to Google Play Apps Categorization.”
- [10] A. Indriani and D. Nbc, “Klasifikasi Data Forum dengan menggunakan Metode Naïve Bayes Classifier,” pp. 5–10, 2014.
- [11] Bustami, “Penerapan Algoritma Naive Bayes,” *J. Inform.*, vol. 8, no. 1, pp. 884–898, 2014.
- [12] S. Rana, “Comparative Analysis of Sentiment Orientation Using SVM and Naïve Bayes Techniques,” no. October, pp. 106–111, 2016.
- [13] Y. K. S. Derick Iskandar, “Perbandingan Akurasi Klasifikasi Tingkat Kemiskinan,” *Perbandingan Akurasi Klasifikasi Tingkat Kemiskin.*, vol. xxx, no. xxx, pp. 1–8, 2014.