

## Analisa Sentimen Kicauan Twitter Tokopedia Dengan Optimalisasi Data Tidak Seimbang Menggunakan Algoritma SMOTE

Andreyestha<sup>1\*</sup>, Qudsiah Nur Azizah<sup>2</sup>

<sup>1,2</sup>Program Studi Sistem Informasi, Universitas Bina Sarana Informatika

\*andreyes2505@bsi.ac.id

### Abstrak

Seiring perkembangan teknologi, media sosial kini sering dimanfaatkan untuk memberikan sebuah penilaian, salah satunya Twitter merupakan salah satu microblogging paling populer. Di Twitter pengguna tidak hanya mereview sebuah produk, tetapi tidak jarang mereka melakukan complain atau berbagi pengalaman terhadap tingkat kepuasan selama menggunakan Tokopedia. Analisa sentimen terhadap tokopeida memberikan indikator yang berguna untuk berbagai tujuan yang dapat ditemukan dalam komentar, umpan balik ataupun kritik. Penelitian ini menggunakan algoritma Naïve Bayes dan Random Forest untuk memberikan hasil klasifikasi yang diharapkan, analisis akan dilakukan dengan membandingkan beberapa kombinasi algoritma yang akan diuji pada kicauan twitter mengenai tokopedia, diantaranya kombinasi dengan Synthetic Minority Oversampling Technique (SMOTE) untuk mengoptimalkan data yang tidak seimbang. Pada pengujian kicauan twitter tokopedia, SMOTE dapat meningkatkan akurasi algoritma Naive Bayes menjadi 86,93% meningkat 3,4% dari sebelumnya 83,53%. Sedangkan Random Forest dengan SMOTE memiliki nilai akurasi sebesar 88,44% meningkat 1,55% dari pengujian Random Forest sebelumnya sebesar 86,89%.

**Kata kunci:** Analisa Sentimen, Data Tidak Seimbang, SMOTE, Text Mining, Tokopedia

### Abstract

Along with the development of technology, social media is now often used to provide an assessment, one of which is Twitter is one of the most popular microblogging. On Twitter, users don't just review a product, but they often complain or share their experiences about their level of satisfaction while using Tokopedia. Sentiment analysis on tokopeida provides useful indicators for various purposes which can be found in comments, feedback or criticism. This study uses the Naïve Bayes and Random Forest algorithms to provide the expected classification results, the analysis will be carried out by comparing several combinations of algorithms that will be tested on twitter tweets about Tokopedia, including the combination with Synthetic Minority Oversampling Technique (SMOTE) to optimize unbalanced data. In the Tokopedia Twitter tweet test, SMOTE was able to increase the accuracy of the Naive Bayes algorithm to 86.93%, an increase of 3.4% from the previous 83.53%. Meanwhile, Random Forest with SMOTE has an accuracy value of 88.44%, an increase of 1.55% from the previous Random Forest test of 86.89%.

**Keywords:** Sentiment Analysis, Unbalanced Data, SMOTE, Text Mining, Tokopedia

### 1. Pendahuluan

Seiring perkembangan teknologi, media sosial kini sering dimanfaatkan untuk memberikan sebuah penilaian atau hanya sekedar menyampaikan keluh kesahnya mengenai hal tertentu. Penilaian

dapat berupa sebuah pujian, kritik, atau bahkan umpatan kekesalan pada sesuatu yang tidak sesuai dengan ekspektasinya. Salah satunya media yang masih ramai penggunaannya hingga

sekarang adalah twitter. Twitter merupakan salah satu microblogging paling populer [1], Penggunaan twitter umumnya diakses para pengguna di Indonesia mencapai 59% dan menduduki peringkat ke-5 media sosial yang sering digunakan pada tahun 2020 [2], menjadikan media sosial tersebut untuk berbagi pemikiran dan pendapat tentang berbagai aspek dan aktivitas, salah satunya mengenai review mereka terhadap suatu e-commerce yaitu Tokopedia. Akibatnya, Twitter dipandang sebagai sumber informasi yang kaya untuk pengambilan keputusan dan analisa sentimen.

Warganet terbiasa mencari informasi sebelum melakukan pembelian di situs belanja online. Informasi dapat dicari berbagai cara tentang suatu produk [3]. Di Twitter pengguna tidak hanya mereview sebuah produk, tetapi tidak jarang mereka melakukan komplain atau berbagi pengalaman terhadap tingkat kepuasan selama menggunakan Tokopedia. Analisa sentimen memberikan indikator yang berguna untuk berbagai tujuan yang dapat ditemukan dalam komentar, umpan balik ataupun kritik.

Penelitian ini menggunakan algoritma Naïve Bayes dan Random Forest untuk memberikan hasil klasifikasi yang diharapkan, analisis akan dilakukan dengan membandingkan beberapa kombinasi algoritma yang akan diuji pada kicauan twitter mengenai tokopedia, diantaranya yaitu Algoritma tersebut akan dikombinasikan dengan

*Synthetic Minority Oversampling Technique* (SMOTE) untuk mengoptimalkan data yang tidak seimbang.

## 2. Tinjauan Pustaka

### 2.1. Penelitian Terkait

Pada penelitian sebelumnya analisa sentimen terhadap kicauan twitter tokopedia mendapat hasil akurasi sebesar 83.33% saat menggunakan algoritma Support Vector Machine (SVM) [4]. Pada penelitian lain analisa sentimen tokopedia mendapat akurasi sebesar 83.97% saat menggunakan algoritma Naive Bayes (NB)[5]. Evaluasi dilakukan dengan menghitung tingkat keakuratan metode analisis opini publik., k-fold cross validation dipilih untuk menguji keakuratan metode Naïve Bayes sehingga diketahui akurasinya. Prinsip cross-validation adalah membagi data menjadi dua bagian data latih dan data uji [6]. Hasil lain juga di dapat saat menggunakan algoritma Decision Tree, KNN, dan Naïve Bayes sebesar 82%, 80%, dan 79% pada analisa sentimen mengenai marketplace seperti tokopedia dan bukalapak [7]. Proses pengujian bertujuan untuk mengetahui akurasi yang diperoleh dari klasifikasi Naive Bayes, metode pengujian yang digunakan adalah pengujian kinerja [8][9].

## 2.2. Landasan Teori

### 1. Text Mining

Text mining adalah proses mengeksplorasi pola-pola yang ada dalam sumber data tekstual untuk menarik informasi dan pengetahuan dari sejumlah besar data. Tujuan dari text mining adalah untuk menganalisis koneksi dokumen menggunakan kata-kata yang mewakili konten untuk mendapatkan informasi yang berguna. Data yang digunakan sebagai input dalam proses text mining tidak terstruktur, sedangkan data mining menggunakan data terstruktur atau database sebagai input. Namun, penambangan teks menggunakan konsep dan teknik penambangan data [10]. Hal ini karena teks diperlukan dalam format yang lebih terstruktur sebelum dapat dilakukan analisis lebih lanjut [11].

### 2. Sentimen Analisis

Sentimen Analisis adalah proses menentukan emosi dan mengklasifikasikan polaritas teks dalam dokumen atau kalimat sehingga kategori dapat diidentifikasi sebagai emosi positif, negatif, atau netral. Analisis sentimen berfokus pada opini positif atau negatif dan juga dapat dibandingkan dengan penambangan opini [12].

### 3. Dataset

Dataset adalah suatu database didalam memori (in-memory). Dataset memiliki semua karakteristik, fitur, dan fungsi dari database biasa. Dataset dapat memiliki banyak tabel-tabel pada

suatu dataset dapat memiliki foreign key dan integritas referensial [13].

Dataset adalah objek yang merepresentasikan data dan relasinya di memory. Strukturnya mirip dengan data yang ada di database. Dataset berisi koleksi dari data tabel dan data.

Jenis dataset ada dua macam yaitu :

- a. Private Dataset yaitu dataset yang dapat diambil dari organisasi yang kita jadikan tempat atau objek penelitian. Adapun contoh-contohnya seperti instansi, rumah sakit, pabrik, perusahaan jasa, etc.
- b. Public Dataset yaitu yang dapat diambil dari repository public yang telah disepakati oleh para peneliti. Adapun contoh-contohnya seperti : UCI.

### 4. Naive Bayes

Naive Bayes adalah salah satu algoritma data mining yang paling populer. Efisiensinya berasal dari asumsi independensi atribut, meskipun ini mungkin dilanggar di banyak kumpulan data dunia nyata. Banyak upaya telah dilakukan untuk mengurangi asumsi tersebut, di antaranya pemilihan atribut merupakan pendekatan yang penting [14].

Naive Bayes merupakan metode baru untuk memprediksi probabilitas. Algoritma ini memanfaatkan teori probabilitas yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan

berdasarkan pengalaman di masa sebelumnya [15].

Dasar dari teorema Bayes dinyatakan dalam Persamaan [16].

$$cNB = \operatorname{argmax}(c) \prod_{x \in X} (x|c)$$

Keterangan:

- $cNB$  = Kelas hasil penghitungan Naïve Bayes.  
 $(c)$  = Probabilitas kelas pada data latih.  
 $(x|c)$  = Peluang fitur  $x$  muncul pada dokumen  $c$ .  
 $x \in X$  = jumlah fitur unik pada dokumen.

Probabilitas parameter  $(c_j)$  dijelaskan pada Persamaan.

$$(c_j) = \frac{N_j}{N}$$

Keterangan:

- $P(c_j)$  = Probabilitas masing-masing kelas pada data latih.  
 $N_j$  = Jumlah dokumen pada suatu kelas pada data latih.  
 $N$  = Jumlah dokumen pada data latih.

### 5. Random Forest

Random Forest adalah salah satu metode ensemble paling kuat dengan kinerja tinggi dalam hal data dimensi tinggi. Dalam pendekatan ini, pepohonan dibatasi pada kedalaman tertentu untuk memungkinkan peningkatan pandangan. Setiap pohon yang dibatasi dianggap sebagai pandangan masalah lokal dan lebih banyak pandangan lokal lebih baik klasifikasinya. Hasil penelitian menunjukkan bahwa dengan mengikat pohon akurasi dapat ditingkatkan untuk masalah dimensi tinggi [17].

### 6. SMOTE

Synthetic Minority Oversampling Technique (SMOTE) merupakan metode oversampling yang digunakan untuk menangani masalah ketidakseimbangan kelas. Teknik ini mensintesis sampel baru dari kelas minoritas untuk menyeimbangkan dataset dengan membuat instance baru dari kelas minoritas dengan membentuk kombinasi cembung dari instance tetangga [18]. SMOTE dibagi menjadi langkah-langkah berikut: de-noising, oversampling, dan penyaringan, bagian ini terutama memperkenalkan proses operasi denoising sampel. Kumpulan data awal dihilangkan noise dan diklasifikasikan sebelumnya oleh algoritma mesin vektor pendukung, untuk menemukan sampel minoritas dengan klasifikasi yang salah dan mengidentifikasi sampel noise. Selanjutnya menentukan kategori sampel tetangga untuk setiap sampel yang salah klasifikasi dan menghilangkan sampel kebisingan dari kelas minoritas.

### 3. Metode Penelitian

Data yang digunakan merupakan kicauan bentuk teks yang dikumpulkan melalui twitter dengan jumlah 328 kicauan dengan kata kunci "tokopedia" yang terdiri dari 200 ulasan positif dan 128 ulasan negatif. Dataset diuji menggunakan metode Naive Bayes Classifier dan Random Forest yang yang di optimalkan menggunakan algoritma Synthetic

Minority Oversampling Technique (SMOTE). Dataset berisi kicauan dengan kata kunci “tokopedia”, data yang di uji berjumlah 328 kicauan dengan rincian 200 ulasan positif dan 128 ulasan negatif, kemudian mengevaluasi hasil akurasi, Precision, Recall dan ROC Area dari algoritma tersebut untuk mengetahui kombinasi pengujian yang tepat menggunakan algoritma Naive Bayes dan Random Forest terhadap analisa sentimen pada kicauan twitter tokopedia yang diharapkan dapat menghasilkan akurasi lebih baik.

### 3.1. Tahapan Proses Text Mining

Dalam text mining terdapat tahapan yang dilakukan, dan tahapan tahapan tersebut saling berkaitan antara satu dengan yang lainnya, berikut untuk penjabarannya [19]:

#### 1. Pengumpulan data

Pada tahap ini dilanjutkan dengan mengumpulkan data yang dibutuhkan atau disebut dengan data awal, data yang didapat merupakan kicauan dengan kata kunci “tokopedia” yang yang didapat melalui twitter.

#### 2. Pre-processing data

Pre-pocessing data mencakup langkah awal untuk mengubah data agar lebih mudah digunakan sebelum analisis sentimen dapat dilakukan. Beberapa tahapan yang dilakukan pada pre-pocessing data adalah mengubah ke dalam bentuk yang lebih informatif.

Tahapan Pre-pocessing data mencakup proses berikut.

#### a. Remove numbers

Angka dihapus dari data karena tidak memainkan peran vital dalam menentukan orientasi teks.

#### b. Remove punctuation

Dalam tes, tanda baca dihilangkan dari teks. Meskipun emoji yang dibuat menggunakan tanda baca dapat membantu dalam memprediksi emosi, emoji bahkan dapat digunakan secara ironis. Penggunaan emoji tidak diperhatikan dan hanya didasarkan pada pengujian berdasarkan analisis kata yang diperoleh dari kalimat.

#### c. Remove stop words

Kumpulan kata seperti kata ganti, kata depan, konjungsi, dll. Kata yang sering muncul dalam data tetapi tidak menyampaikan informasi yang berarti atau penting tentang nilai sentimental dari sebuah kalimat disebut stopword. Menghilangkan stopword dapat membantu mengurangi kebutuhan memori saat menyortir ulasan.

#### d. Strip whitespaces

Whitespaces tidak memiliki tujuan yang berguna saat digunakan sehingga dihapus dari teks aslinya.

e. Convert to lower case

Konsistensi perlu dijaga dan memetakan kata-kata terlepas dari kasusnya, kalimat akan dikonversi menjadi huruf kecil.

f. Stemming

Banyak kata yang berasal dari kata dasar, setiap kata dihilangkan awalan dan akhirnya. Stemming dapat secara signifikan mengurangi beban memori selama pelatihan dan klasifikasi.

### 3. Pengolahan Data

Pengolahan data dilakukan dengan menggunakan algoritma data mining termasuk akurasi atau biasa disebut dengan seleksi fitur dan ekstraksi menggunakan algoritma data mining. Pada titik ini, pertama lakukan seleksi fitur sebelum mengekstrak fitur. Teknik yang paling penting dalam preprocessing adalah proses seleksi fitur. Teknik ini mengurangi jumlah fitur yang terlibat dalam menentukan nilai kelas target, yang pengambilannya data terstruktur. Data diperoleh dan dilakukan seleksi kemudian data tersebut diekstraksi. Proses ini merupakan cara untuk mengidentifikasi suatu objek. Langkah ini akan mencocokkan kepribadian dengan data yang diperoleh.

### 4. Hasil dan Pembahasan

Penelitian didukung dengan tools WEKA 3.8.1. Penelitian ini menggunakan dataset kicauan dari

twitter mengenai tokopedia berjumlah 328 ulasan yang terdiri dari 200 ulasan positif dan 128 ulasan negatif. Pengujian dilakukan dengan menguji coba model usulan terhadap dataset. Model usulan yang diuji meliputi Naïve Bayes (NB), Random Forest (RF) dan optimalisasi dataset tidak seimbang dengan menggunakan *Synthetic Minority Oversampling Technique* (SMOTE). Pengujian terhadap algoritma tersebut dilakukan secara bergantian.

Pertama, pengujian dilakukan terhadap model Naïve Bayes. Hasil pengujian dihitung dengan menggunakan Confusion Matrix untuk mencari akurasi, sensitivitas/recall/TPrate, FPrate, Precision/PPV, F-Measure dan AUC. Hasil perhitungan diperoleh probabilitas sebagai berikut.

Tabel 1. Hasil Pengujian pada algoritma Naive Bayes

Perhitungan	Persentase
Akurasi	83,53%
Sensitivitas/Recall/TPrate	0,835
Fprate	0,164
Precision/PPV	0,840
F-Measure	0,837
Area under ROC/ AUC	0,887

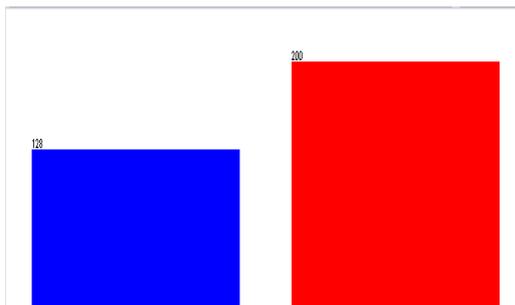
Kedua, pengujian dilakukan terhadap model Random Forest. Hasil pengujian dihitung dengan menggunakan Confusion Matrix untuk mencari akurasi, sensitivitas/recall/TPrate, FPrate, Precision/PPV, F-Measure, dan AUC. Hasil

perhitungan diperoleh probabilitas sebagai berikut.

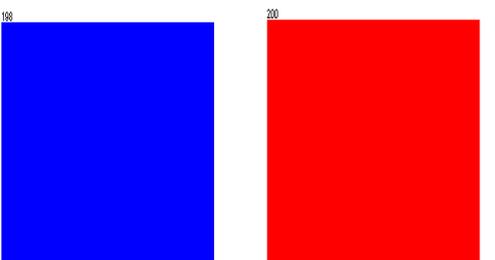
Tabel 2. Hasil Pengujian pada algoritma Random Forest

Perhitungan	Persentase
Akurasi	86,89%
Sensitivitas/Recall/TPrate	0,869
Fprate	0,194
Precision/PPV	0,882
F-Measure	0,864
Area under ROC/ AUC	0,954

Ketiga, penerapan SMOTE meminimalisasi ketidak seimbangan kelas pada dataset kicauan tokopedia, sehingga total keseluruhan data terdiri dari 398 kicauan, yaitu 200 merupakan kicauan positif dan 198 merupakan kicauan negatif.



Gambar 1 Dataset Ketidak Seimbangan Awal



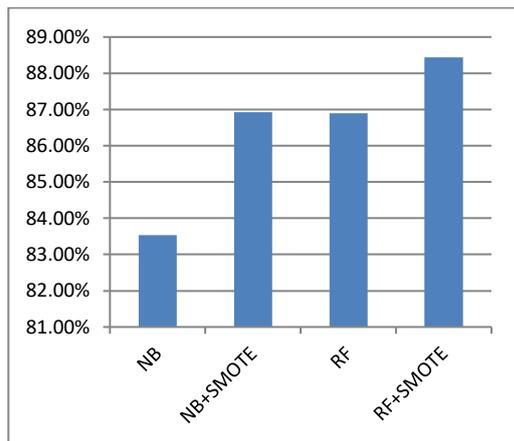
Gambar 2 Dataset Ketidak Seimbangan Akhir dengan SMOTE

Hasil pengujian dihitung dengan menggunakan Confusion Matrix untuk mencari akurasi, sensitivitas/recall/TPrate, FPrate, Precision/PPV, F-Measure, dan AUC. Hasil perhitungan diperoleh probabilitas sebagai berikut.

Tabel 3. Hasil Pengujian Optimalisasi SMOTE

Perhitungan	SMOTE+ NB	SMOTE+ RF
Akurasi	86,93%	88,44%
Sensitivitas/Recall/TPrate	0,869	0,884
Fprate	0,131	0,117
Precision/PPV	0,869	0,899
F-Measure	0,869	0,883
Area under ROC/ AUC	0,920	0,962

Dari hasil klasifikasi dengan mengintegrasikan SMOTE dari kicauan dengan kata kunci tokopedia, diketahui SMOTE dapat meningkatkan hasil akurasi dari algoritma Naïve Bayes dan Random Forest. Naive Bayes dengan SMOTE memiliki nilai akurasi sebesar 86,93% meningkat 3,4% dari pengujian Naive Bayes sebelumnya sebesar 83,53%. Sedangkan Random Forest dengan SMOTE memiliki nilai akurasi sebesar 88,44% meningkat 1,55% dari pengujian Random Forest sebelumnya sebesar 86,89%.



Gambar 3 Grafik Perbandingan Akurasi

## 5. Kesimpulan

Penelitian ini mencoba mengoptimalkan metode Naive Bayes dan Random Forest yang dikombinasikan dengan algoritma Synthetic Minority Oversampling Technique (SMOTE), untuk memberikan hasil akurasi tertinggi pada dataset kicauan twitter tokopedia yang tidak seimbang.

Dalam hasil pengujian yang dilakukan sebelum dioptimalkan dengan algoritma SMOTE, metode Naive Bayes mendapat akurasi 83,53% dan metode Random Forest mendapat akurasi 86,89%. Sedangkan pengujian dengan mengoptimalkan metode Naive Bayes dan Random Forest menggunakan algoritma SMOTE memiliki hasil akurasi lebih baik, dimana Naive Bayes mendapatkan akurasi 86,93% dan Random Forest mendapat akurasi 88,44%.

## 6. Daftar Pustaka

- [1] D. Suparto, "Efektivitas Penggunaan Sosial Media Twitter Dalam Penyebaran Informasi Dalam Pelayanan Publik (Studi Kasus Kabupaten Pemalang)," *Indones. Gov. J. (Kaji. Polit. – Pemerintah. )*, vol. 04, no. 02, pp. 161–172, 2021.
- [2] T. Krisdiyanto, "Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 7, no. 1, pp. 32–37, 2021.
- [3] P. Arbaini, "Pengaruh Consumer Online Rating Dan Review Terhadap Keputusan Pembelian Pada Pengguna Marketplace Tokopedia," *J. Bisnis dan Manaj.*, vol. 7, no. 1, pp. 25–33, 2020, doi: 10.26905/jbm.v7i1.3897.
- [4] S. Pada, U. Produk, and T. Menggunakan, "Media Informatika Vol.20 No.2 (2021) 97," vol. 20, no. 2, pp. 97–108, 2021.
- [5] K. Ramadhan and K. M. L, "Analisis Sentimen Terhadap Toko Online Menggunakan Naïve Bayes pada Media Sosial Twitter," *e-Proceeding Eng.*, vol. 5, no. 3, pp. 8141–8151, 2018.
- [6] D. P. Artanti, A. Syukur, A. Prihandono, and D. R. I. M. Setiadi, "Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes," pp. 8–9, 2018.
- [7] I. Saputra, H. E. Darono, F. Amsury, M. R. Fahdia, and B. Ramadhan, "Analisis Sentimen Pengguna Marketplace Bukalapak dan Tokopedia di Twitter Menggunakan Machine Learning," *Fakt. Exacta*, vol. 13, no. 4, pp. 200–207, 2020, doi: 10.30998/faktorexacta.v13i4.7074.
- [8] A. Sri Widagdo, B. S. W.A, and A. Nasiri, "Analisis Tingkat Kepopuleran E-Commerce Di Indonesia Berdasarkan Sentimen Sosial Media Menggunakan Metode Naïve Bayes," *J. Inf.*, vol. 6, pp. 1–5, 2020.
- [9] M. Wasil, A. Sudianto, and Fathurrahman,

- "Application of the Decision Tree Method to Predict Student Achievement Viewed from Final Semester Values," *J. Phys. Conf. Ser.*, vol. 1539, no. 1, 2020, doi: 10.1088/1742-6596/1539/1/012027.
- [10] A. Sudianto and M. Wasil, "Penerapan Sistem Informasi Geografis dalam Pemetaan Sebaran Kasus Gizi Buruk Lombok Timur merupakan salah satu Kabupaten yang berada di Provinsi Nusa Tenggara Barat yang terletak di sebelah Timur Pulau Lombok, Kabupaten Lombok Timur Gizi Buruk Malnutrisi," vol. 4, no. 2, pp. 142–150, 2021.
- [11] A. F. Akbar, "User Perception Analysis of Online Learning Platform ' Zenius ' During the Coronavirus Pandemic Using Text Mining Techniques," vol. 17, no. 2, pp. 33–47, 2021.
- [12] Samsir, Ambiyar, U. Verawardina, F. Edi, and R. Watrionthos, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 149, 2021, doi: 10.30865/mib.v5i1.2604.
- [13] M. H. N. Yahya, "Penggunaan Algoritma Support Vector Machine (SVM) Untuk Penentuan Kelayakan Pemberian Kredit Koperasi Serba Usaha 'Daruzzakah Zakah' Desa Rensing Kecamatan Sakra Lombok Timur Nusa Tenggara Barat," *Infotek J. Inform. dan Teknol.*, vol. 3, no. 1, pp. 32–41, 2020.
- [14] S. Chen, G. I. Webb, L. Liu, and X. Ma, "A novel selective naïve Bayes algorithm," *Knowledge-Based Syst.*, vol. 192, no. xxx, p. 105361, 2020, doi: 10.1016/j.knosys.2019.105361.
- [15] M. S. Yupi Kuspani Putra, "Perbandingan Algoritma Naive Bayes dan Naive Bayes Berbasis PSO untuk Analisis Kredit pada PT. BPR Syariah Paokmotong," *Infotek J. Inform. dan Teknol.*, vol. 2, no. 2, pp. 61–69, 2019.
- [16] N. D. Pratama, Y. A. Sari, and P. P. Adikara, "Analisis Sentimen Pada Review Konsumen Menggunakan Metode Naive Bayes Dengan Seleksi Fitur Chi Square Untuk Rekomendasi Lokasi Makanan Tradisional," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 9, pp. 2982–2988, 2018.
- [17] K. Shah, H. Patel, D. Sanghvi, and M. Shah, "A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification," *Augment. Hum. Res.*, vol. 5, no. 1, 2020, doi: 10.1007/s41133-020-00032-0.
- [18] Hermanto, A. Y. Kuntoro, T. Asra, E. B. Pratama, L. Effendi, and R. Ocanitra, "Gojek and Grab User Sentiment Analysis on Google Play Using Naive Bayes Algorithm and Support Vector Machine Based Smote Technique," *J. Phys. Conf. Ser.*, vol. 1641, no. 1, 2020, doi: 10.1088/1742-6596/1641/1/012102.
- [19] M. P. R. Putra and K. R. N. Wardani, "Penerapan Text Mining Dalam Menganalisis Kepribadian Pengguna Media Sosial," *JUTIM (Jurnal Tek. Inform. Musirawas)*, vol. 5, no. 1, pp. 63–71, 2020, doi: 10.32767/jutim.v5i1.791.