SNESTIK

Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika

https://ejurnal.itats.ac.id/snestik dan https://snestik.itats.ac.id

Informasi Pelaksanaan :
SNESTIK V - Surabaya, 26 April 2025
Fakultas Teknik Elektro dan Teknologi Informasi, Institut Teknologi Adhi Tama Surabaya

Informasi Artikel:
DOI : 10.31284/p.snestik.2025.7222

Prosiding ISSN 2775-5126

Fakultas Teknik Elektro dan Teknologi Informasi-Institut Teknologi Adhi Tama Surabaya
Gedung A-ITATS, Jl. Arief Rachman Hakim 100 Surabaya 60117 Telp. (031) 5945043
Email : snestik@itats.ac.id

Pendekatan Ekstraksi Informasi Kualitas Layanan Toko di Platform Shopee Berdasarkan Komentar Menggunakan Ensemble Learning

Achmad Maududie[1], Dini Aulia Hidayati[2], dan Harry Soepandi[3]
Universitas Jember[1], Universitas Jember [2], Universitas Jember [3]
e-mail: maududie@unej.ac.id

ABSTRACT

The service quality of a shop can be reviewed through the rating feature which is represented in the form of the number of stars and comments. Customers often provide ratings and comments after purchasing products and receiving service from a shop. However, this is inefficient because potential buyers have to look at the comments one by one. Therefore, Machine Learning is one way to classify these comments. Creating a Service Quality Dictionary is an important point in this classification process because it functions to determine the service class of each comment. The algorithms used include Support Vector Machine, Random Forest, Multinomial Naive Bayes, TF-RF, and Ensemble Learning Bagging. The classification stages include the process of scraping, labeling, data sharing, preprocessing, word weighting, modeling, and model evaluation. Data is classified into two labels, the Service Quality label and the Not Service Quality label. The Service Quality Label will then be reclassified into three labels, namely Positive, Negative and Neutral. Modeling evaluation uses training data with three ratios, namely 90:10, 80:20, and 70:30. From these three ratios, it was found that the 90:10 ratio had better accuracy results, so the modeling evaluation for testing data used the ratio that had the best results from the training data evaluation, namely 90:10 which produced an accuracy of 87,3%.

Keywords: *Ensemble learning bagging; service quality; text mining; TF-RF.*

ABSTRAK

Kualitas layanan dari sebuah toko dapat ditinjau melalui fitur rating yang direpresentasikan dalam bentuk jumlah bintang dan komentar. Pelanggan sering memberikan rating dan komentar setelah membeli produk dan mendapatkan layanan dari sebuah toko. Namun hal ini menjadi tidak efisien karena calon pembeli harus melihat satu persatu komentar tersebut. Maka dari itu, Machine Laerning menjadi salah satu cara dalam mengklasifikan komentar-komentar tersebut. Pembuatan Kamus Kata Kualitas Layanan menjadi poin

penting dalam proses klasifikasi ini karena berfungsi untuk menentukan kelas layanan dari masing-masing komentar. Algoritma yang digunakan meliputi Support Vector Machine, Random Forest, Multinomial Naive Bayes, TF-RF, serta Ensemble Learning Bagging. Tahapan klasifikasi meliputi proses scraping, pelabelan, pembagian data, preprocessing, pembobotan kata, pemodelan, dan evaluasi model. Data diklasifikasikan menjadi dua label, label Kualitas Layanan dan label Bukan Kualitas Layanan. Setelah melalui proses filtering, Label Kualitas Layanan kemudian akan diklasifikasikan kembali menjadi tiga label, yaitu Positif, Negatif, dan Netral. Evaluasi pemodelan menggunakan data training dengan tiga rasio, yaitu 90:10, 80:20, dan 70:30. Dari ketiga rasio tersebut didapatkan bahwa rasio 90:10 memiliki mayoritas hasil akurasi yang lebih baik, sehingga evaluasi pemodelan untuk data testing menggunakan rasio yang memiliki hasil paling baik dari evaluasi data training, yaitu 90:10 yang menghasilkan akurasi 87,3%.
Kata kunci: ensemble learning bagging, kualitas layanan, text mining, metode pembobotan kata TF-RF.

PENDAHULUAN

Menurut Cronin [2], kualitas layanan menjadi faktor kunci dalam mendorong niat beli. Kualitas layanan dapat mempengaruhi loyalitas pembeli secara langsung dan tidak langsung [3]. Kualitas layanan dari sebuah toko dapat ditinjau melalui fitur rating yang direpresentasikan dalam bentuk jumlah bintang dan komentar. Namun demikian, hal ini menjadi kurang efisien karena pembeli harus mengecek dan mengklasifikasi satu persatu komentar pembeli sebelumnya terhadap kualitas layanan yang diberikan oleh toko tersebut. Hanya dengan melihat jumlah bintang saja dirasa tidak cukup untuk sepenuhnya menggambarkan isi dari sebuah komentar yang bisa memuat lebih dari satu penilaian atribut produk [4].

Salah satu cara yang dapat digunakan dalam pengklasifikasian adalah machine learning. Machine Learning merupakan sistem pembelajaran yang mampu mempelajari pola dan informasi secara otomatis tanpa memerlukan pemrograman berulang oleh manusia [5]. Supervised machine learning merupakan salah satu teknik dalam machine learning yang digunakan untuk mengidentifikasi hubungan antara atribut input dan atribut target. Salah satu bentuk supervised learning adalah klasifikasi [6]. Menurut Septianto dalam [7], klasifikasi adalah proses mengelompokkan data atau objek ke dalam kelas atau kategori tertentu berdasarkan variabel yang telah ditentukan. Dengan demikian teknik supervised machine learning ini sangat mungkin digunakan untuk mengklasifikasi komentar pembeli berdasarkan opini (sentimen) atas kualitas layanan di setiap toko, sehingga pengguna tidak perlu lagi meninjau satu persatu komentar pembeli sebelumnya.

Melihat permasalahan di atas serta keunggulan ensemble learning, maka penelitian ini mencoba memanfaatkan metode klasifikasi dalam machine learning menggunakan teknik ensemble learning yang melibatkan tiga algoritma, yaitu Support Vector Machine, Random Forest, dan Mutinomial Naïve Bayes untuk mengelompokkan komentar pembeli dari masing-masing toko dalam platform Shopee. Komentar tersebut dikelompokkan berdasarkan sentimen atas layanan yang telah diberikan dan berdasarkan informasi tersebut akan disimpulkan bagaimana kualitas layanan yang diberikan toko-toko tersebut.

METODE

Tahapan Penelitian

Tahapan penelitian yang akan dilakukan pada penelitian ini dijabarkan pada Gambar 1.

Pembuatan Kamus Kata Kualitas Layanan

Kamus layanan kata digunakan untuk acuan pelabelan. Adapun kamus layanan kata ini merujuk pada indikator kualitas layanan yang meliputi [8]:

1. *Fulfilment/Reability*, meliputi dua hal yaitu (1) keakuratan tampilan dan deskripsi produk sehingga produk yang diterima pembeli sesuai dengan ekspektasi mereka saat melihat dan membaca deskripsi produk dan (2) pengiriman produk yang sesuai dengan waktu yang telah dijanjikan. Tabel 1 merupakan contoh komentar yang merepresentasikan *fulfilment/reability*.

Tabel 1. Contoh komentar yang mengandung kata pada indikator 1.

Positif	Negatif	Netral
barang sesuai pesanan. pengiriman cepat. terima kasih penjual	pengiriman dan kemasan lama sekali	Berfungsi dengan baik. Pengiriman dan pengemasan oke

2. *Layanan* pelanggan atau *customer service* yang tanggap, membantu, dan siap untuk merespons pertanyaan pelanggan dengan cepat. Tabel 2 merupakan contoh komentar yang merepresentasikan *customer service*.

Tabel 2. Contoh komentar yang mengandung kata pada indikator 2.

Positif	Negatif	Netral
Barang bagus, respons admin baik ramah makasih penjual	udah chat admin cm gada respons solusinya gmn	harga cukup bersaing, respons penjual lumayan

3. Kompensasi penjual yang merujuk pada ganti rugi kepada konsumen apabila ada barang *yang* kurang memuaskan. Contoh komentar yang merepresentasikan kompensasi penjual dapat dilihat pada Tabel 3.

Tabel 3. Contoh komentar yang mengandung kata pada indikator 3.

Positif	Negatif	Netral
respons untuk komplainnya cepat dan jelas	gabisa nyala charge gabisa minta ganti barang/dana	Udah coba tapi tetep gak bisa dipake. Udh chat penjual juga

Gambar 1. Tahapan Penelitian

Scraping Data

Menurut Turland [9] *Web scraping* adalah proses ekstraksi dokumen semi-terstruktur dari internet, khususnya halaman web yang umumnya ditulis dalam bahasa *markup* seperti *HTML* atau *XHTML*. *Scraping data* atau pengambilan data dilakukan menggunakan kode program yang dibuat dengan bahasa pemrograman *Python*. Spesifikasi data yang diambil merupakan *review* pembeli yang terdapat pada fitur *rating*.

Pelabelan Data

Pelabelan data dilakukan dengan tujuan untuk mengelompokkan data ke dalam beberapa kelas sentimen. Pada analisis sentimen, umumnya ada tiga label yang digunakan, yaitu positif, negatif, dan netral. Pemberian label dilakukan dalam dua tahap, yaitu pemberian label layanan dan pemberian label sentimen. Pemberian label layanan meliputi dua kategori, yaitu komentar tentang kualitas layanan dan komentar bukan tentang kualitas layanan. Tabel 4 menunjukkan contoh label layanan.

Tabel 4. Contoh label layanan.

Komentar	Label
Next beli lagi ditoko ini pelayanan dan pengirimannya cepet dan barangnya good banget meskipun murce tapi no kaleng kaleng	Kualitas Layanan
bagus banget pancinya serbaguna bisa dipake masak apa aja keren	Bukan Kualitas Layanan

Sedangkan pemberian label sentimen meliputi tiga kategori, yaitu “Baik”, “Tidak Baik”, dan “Netral”. Contoh label sentimen bisa dilihat melalui Tabel 5.

Tabel 5. Contoh label sentimen.

Komentar	Label
Barang sesuai deskripsi, pengiriman cepat, harga bagus, respons cepat.	Positif
JELEK SEKALI! UDAH PATAH JANGAN dibeli JELEK bgt? PATAH kardus UDH PADA PENYOK JGN dibeli!	Negatif
Packingan standart estimasi pengiriman juga standart	Netral

Preprocessing

Menurut [10], dalam teks mining, preprocessing adalah proses yang bertanggung jawab mengubah data tidak terstruktur yang tersimpan dalam kumpulan dokumen menjadi format yang lebih sesuai, jelas dan terstruktur. Terdapat enam tahap yang akan dilakukan pada penelitian ini, yaitu:

1. *Cleaning*

   Tahap cleaning digunakan untuk membersihkan data teks dari karakter selain huruf seperti tanda baca, simbol, atau karakter lain yang tidak relevan dengan analisis teks. Tabel 6 merupakan contoh hasil tahap cleansing.

   Tabel 6. Contoh hasil tahap cleansing.

Sebelum	Sesudah
proses Packing dan pengirimannya cepet banget ! ?? pancinya juga cepet panas . semoga barangnya awet	proses Packing dan pengirimannya cepet banget pancinya juga cepet panas semoga barangnya awet

2. *Case Folding*

   Tahap case folding bertujuan untuk menyeragamkan karakter pada data. Umumnya terdapat dua jenis case folding yang dapat digunakan, yaitu merubah seluruh huruf menjadi huruf kapital (uppercase) atau menjadi huruf kecil (lowercase). Tabel 7 merupakan contoh hasil tahap case folding.

Tabel 7. Contoh hasil tahap case folding.

Sebelum	Sesudah
proses Packing dan pengirimannya cepet banget pancinya juga cepet panas semoga barangnya awet	proses packing dan pengirimannya cepet banget pancinya juga cepet panas semoga barangnya awet

3. Normalisasi
Tahap ini merupakan tahap normalisasi data teks yang mengandung singkatan, bahasa asing, atau kata slang. Tabel 8 merupakan contoh hasil tahap normalisasi.

Tabel 8. Contoh hasil tahap normalisasi.

Sebelum	Sesudah
proses packing dan pengirimannya cepet banget pancinya juga cepet panas semoga barangnya awet	proses kemasan dan pengirimannya cepat sekali pancinya juga cepat panas semoga barangnya tahan lama

4. Tokenizing
Tokenisasi merupakan sebuah proses untuk membagi kalimat atau dokumen menjadi kata-kata individual atau token kata. Tabel 10 merupakan contoh hasil tahap tokenizing.

Tabel 9. Contoh hasil tahap tokenizing.

Sebelum	Sesudah
proses kemasan dan pengirimannya cepat sekali pancinya juga cepat panas semoga barangnya tahan lama	['proses', 'kemasan', 'dan', 'pengirimannya', 'cepat', 'sekali', 'pancinya', 'juga', 'cepat', 'panas', 'semoga', 'barangnya', ‘tahan’, ‘lama]

5. Stopword Removal
Tahap *stopword* removal bertujuan untuk menghapus kata umum yang sering muncul namun tidak memberikan informasi penting, seperti dan, atau, yang, di, dan sebagainya. Tabel 11 merupakan contoh hasil tahap stopword removal.

Tabel 40. Contoh hasil tahap stopword removal.

Sebelum	Sesudah
['proses', 'kemasan', 'dan', 'pengirimannya', 'cepat', 'sekali', 'pancinya', 'juga', 'cepat', 'panas', 'semoga', 'barangnya', ‘tahan’, ‘lama]	['proses', 'kemasan', 'pengirimannya', 'cepat', 'sekali', 'pancinya', 'cepat', 'panas', 'semoga', ‘barangnya’, ‘tahan’, ‘lama]

6. Stemming
Tahap stemming merupakan proses menghilangkan imbuhan pada kalimat untuk mendapatkan kata dasar dari masing-masing kata. Tabel 12 merupakan contoh hasil tahap stemming.

Tabel 51. Contoh hasil tahap stemming.

Sebelum	Sesudah
['proses', 'kemasan', 'pengirimannya', 'cepat', 'sekali', 'pancinya', 'cepat', 'panas', 'semoga', ‘barangnya’, ‘tahan’, ‘lama]	['proses', 'kemas', 'kirim', 'cepat', 'sekali', 'panci', 'cepat', 'panas', 'moga', ‘barang’, ‘tahan’, ‘lama]

Pembobotan Kata

Pembobotan kata merupakan tahap memberi nilai atau bobot pada kata dalam sebuah teks [11]. TF-RF merupakan sebuah teknik pembobotan yang relatif baru, dirancang sebagai usaha untuk meningkatkan metode-metode yang telah ada sebelumnya TF-RF merupakan salah satu metode yang memperhatikan istilah atau kata yang sering muncul dalam setiap dokumen [12]. Perhitungan TF-RF dapat dilihat dari persamaan (1) [13].

$$tf_{tf}rf = tf_{td} \times \log\left(2 + \frac{b}{\max(1, c)}\right) \quad (1)$$

Keterangan :

$tf_{tf}.rf$: nilai bobot kata
tf_{td} : frekuensi kemunculan kata t dalam dokumen
b : banyaknya dokumen yang mengandung kata t
c : jumlah dokumen yang tidak mengandung kata t

Berdasarkan hasil perhitungan di atas, didapatkan hasil perhitungan TF-RF seperti pada Tabel 12.

Tabel 12. Contoh hasil perhitungan TF-RF

Komentar	Kata	TF	RF	TFRF
['barang', 'bagus', 'harga', 'jangkau', 'kirim', 'cepat'	Barang	1178	1,1464	1.350,46
	Bagus	1489	1,1425	1.701,18
	Harga	1332	1,1255	1.499,17
	Jangkau	32	1,0986	35,1552
	Kirim	1536	1,1272	1.731,38
	Cepat	1080	1,1302	1.220,62

Ensemble Learning

Ensemble Learning adalah kombinasi dari beberapa algoritma dengan tujuan untuk mencapai hasil yang lebih baik [14]. Teknik ensemble learning yang bisa digunakan, yaitu [15] bagging, boosting, stacking, dan voting.

Support Vector Machine

Support Vector Machine (SVM) adalah salah satu metode Machine Learning yang digunakan untuk tugas klasifikasi data. Konsep SVM dapat diartikan sebagai upaya menemukan hyperplane terbaik yang berfungsi sebagai pemisah antara dua kelas pada ruang input [16].

Random Forest

Random Forest terdiri dari sejumlah pohon keputusan (decision trees) sebagai model dasar. Setiap pohon keputusan melakukan bootstrap untuk mengambil sampel secara acak. Kemudian melakukan prediksi berdasarkan hasil dari semua pohon keputusan menggunakan majority vote.

Multinomial Naive Bayes

Multinomial Naïve Bayes bekerja berdasarkan konsep term frequency, yang mengindikasikan seberapa sering suatu kata muncul dalam sebuah dokumen. Model ini memiliki kemampuan untuk mengklasifikasikan data yang tidak dapat direpresentasikan secara numerik dengan memperhitungkan frekuensi kata-kata dalam kategori atau kelas yang diberikan [17]. Perhitungan Multinomial Naïve Bayes dilakukan dengan cara menghitung jumlah kemunculan kata dalam dokumen dengan rumus [18]:

$$P(X|c) = log\frac{N_c}{N} + \sum_{i=1}^{n} log\frac{t_i + \alpha}{\sum_{i=1}^{n} t_i + \alpha} \quad (2)$$

Evaluasi Model

Confusion matrix merupakan model data untuk mengevaluasi seberapa baik kinerja model dalam melakukan klasifikasi. Penggunaan *confusion matrix* membantu untuk menghitung evaluasi model seperti *accuracy*, *precision*, *recall*, dan *F1-Score*.

HASIL DAN PEMBAHASAN

Pemodelan

Pada tahapan ini dibuatlah model dari data latih yang nantinya diimplementasikan pada data uji. Terdapat dua model yang dihasilkan, yaitu model klasifikasi layanan dan model klasifikasi sentimen. Pemodelan dilakukan menggunakan tiga algoritma klasifikasi, yaitu Support Vector Machine, Random Forest, dan Multinomial Naive Bayes.

Evaluasi Model Layanan dan Model Sentimen

Evaluasi model layanan dan model sentimen dilakukan dengan menggunakan confusion matrix untuk menghitung nilai accuracy, precision, recall, dan F1-Score. Evaluasi pemodelan ini melibatkan tiga rasio seperti yang telah direncanakan (90:10, 80:20, dan 70:30). Hasil dari evaluasi model layanan dan model sentimen dapat ditinjau pada Tabel 13.

Tabel 63. Hasil evaluasi data testing

Rasio	Evaluasi Model Layanan			Evaluasi Model Sentimen		
	SVM	RF	MNB	SVM	RF	MNB
90 : 10	92,5%	93,2%	89,5%	89,6%	91,2%	82,5%
80 : 20	91,8%	93%	89%	88,8%	91,6%	81,6%
70 : 30	92,5%	92,6%	87,8%	87%	90,7%	81,1%

Berdasarkan hasil dari evaluasi tersebut, maka model yang digunakan untuk memprediksi sentimen untuk data komentar adalah model yang dibangkitkan dari rasio 90:10. Adapun hasil evaluasi model ensemble melalui pendekatan soft voting didapatkan hasil akurasi dan F1-*score* sebesar 87,3%.

Hasil Implementasi Model

Hasil yang didapatkan dari pemodelan klasifikasi layanan ini adalah dataset terbagi menjadi dua label, yaitu kualitas layanan dan bukan kualitas layanan. Tabel 14 menjabarkan hasil dari klasifikasi komentar.

Tabel 74. Hasil klasifikasi komentar

Komentar	Prediksi Label
barangnya udah sampai dengan selamat, Dari Packingannya rapi dan penjualnya ramah	1
simpel dan praktis cocok bgt buat olahraga nyerep keringat	0

Dari 27.223 data yang digunakan, didapatkan 12.931 data berlabel 1 (kualitas layanan) dan 14.292 data berlabel 0 (bukan kualitas layanan). Data dari hasil implementasi klasifikasi layanan kemudian di filtering untuk memisahkan data yang berlabel kualitas layanan dan bukan kualitas layanan. Data yang berlabel kualitas layanan diambil untuk selanjutnya digunakan sebagai dataset dalam mengimplementasikan model sentimen. Sebanyak 12.931 data akan di klasifikasikan menjadi 3 label, yaitu positif, negatif, dan netral. Hasil implementasi pemodelan klasifikasi sentimen bisa dilihat pada Tabel 15 berikut.

Tabel 85. Hasil klasifikasi sentimen

Komentar	Prediksi Label
Ini panasnya cepet banget tadi dihidupkan terus ditinggal ambil bahan udah panas minyaknya pengirimannya cepet langsung dikirim semoga awet deh	1
Sayang pelayanannya perlu ditingkatkan untuk jasa pengiriman, hampir 3 minggu baru nyampe	-1
pengiriman oke Packing oke.. Next update buat masak dan tumis. baru nyoba masak indomie, dan overall good.	0

Dari total 12.931 data kualitas layanan, didapatkan 7.962 data berlabel positif dengan persentase 61%, 2.511 data berlabel negatif dengan persentase 20%, dan 2.458 data berlabel netral dengan persentase 19%. Jumlah data dari masing-masing label dapat dilihat pada tabel 18.

Tabel 9. Jumlah data masing-masing label.

Label	Jumlah	Persentase
Positif	7.962	61%
Negatif	2.511	20%
Netral	2.458	19%

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa; proses klasifikasi kualitas layanan toko di platform Shopee dilakukan dalam dua belas tahap yang meliputi, membuat kamus kata kualitas layanan, scraping data komentar, pelabelan data, preprocessing, pembobotan kata, membagi dataset, pembuatan model, evaluasi model klasifikasi, implementasi model klasifikasi layanan, filtering komentar, implementasi model klasifikasi sentimen, dan menarik kesimpulan dari opini pembeli; dan berdasarkan hasil evaluasi model, didapatkan kinerja yang cukup baik yaitu dengan nilai akurasi dan F1-score sebesar 87,3%. Dari contoh kasus analisis kualitas layanan toko di platform Shopee, model yang dihasilkan dapat disimpulkan bahwa dari 10 toko yang dipilih memiliki kualitas layanan yang baik. Hal ini ditunjukkan bahwa dari 12.931 data berlabel Kualitas Layanan, 7.962 data berlabel positif dengan persentase 61%, 2.511 data berlabel negatif dengan persentase 20%, dan 2.458 data berlabel netral dengan persentase 19%.

DAFTAR PUSTAKA

[1] Databoks, "Pengunjung Shopee Makin Banyak, Bagaimana E-Commerce Lain?," Databoks. Accessed: Dec. 30, 2023. [Online]. Available: https://databoks.katadata.co.id/datapublish/2023/10/11/pengunjung-shopee-makin-banyak-bagaimana-e-commerce-lain

[2] B. C. Adyanto and S. B. Santosa, "Pengaruh kualitas layanan, brand image, harga dan kepercayaan produk terhadap keputusan pembelian (studi layanan e-commerce Berrybenka. com)," *Diponegoro J. Manag.*, pp. 10–29, 2018.

[3] D. Aryani and F. Rosinta, "Pengaruh kualitas layanan terhadap kepuasan pelanggan dalam membentuk loyalitas pelanggan," *BISNIS BIROKRASI J. Ilmu Adm. dan Organ.*, 2010.

[4] E. H. Muktafin, K. Kusrini, and E. T. Luthfi, "Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing," *J. Eksplora Inform.*, vol. 10, no. 1, pp. 32–42, Sep. 2020, doi: 10.30864/eksplora.v10i1.390.

[5] I. Cholissodin, S. Sutrisno, A. A. Soebroto, U. Hasanah, and Y. I. Febiola, *AI, Machine Learning & Deep Learning*. 2020.

[6] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, "Analisis klasifikasi sentimen terhadap sekolah daring pada twitter menggunakan Supervised Machine Learning," *J. Tek. Inform. dan Sist. Inf.*, 2021.

[7] A. C. Imanda, N. Hidayat, and M. T. Furqon, "Klasifikasi Kelompok Varietas Unggul Padi Menggunakan Modified K-Nearest Neighbor," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, pp. 2392–2399, 2018.

[8] M. Wolfinbarger and M. C. Gilly, "eTailQ: Dimensionalizing, measuring and predicting etail quality," *J. Retail.*, vol. 79, no. 3, pp. 183–198, 2003, doi: 10.1016/S0022-4359(03)00034-4.

[9] A. Josi and L. Andretti Abdillah, "PENERAPAN TEKNIK WEB SCRAPING PADA MESIN PENCARI ARTIKEL ILMIAH," 2014.

[10] R. Feldman and J. Sanger, *The text mining handbook : advanced approaches in analyzing unstructured data*. Cambridge University Press, 2007.

[11] M. R. Faisal, D. Kartini, A. R. Arrahimi, and T. H. Saragih, *Belajar Data Science: Text Mining Untuk Pemula I*. 2022.

[12] M. S. Al Ghofany, R. Dwiyansaputra, F. Bimantoro, and Khairunnas, *Indonesian SMS Spam Detection Using TF-RF Feature Weighting Method and Support Vector Machine Classifier*. 2023.

[13] R. N. Aufa, S. S. Prasetiyowati, and Y. Sibaroni, "The Effect of Feature Weighting on Sentiment Analysis TikTok Application Using The RNN Classification," *Build. Informatics, Technol. Sci.*, vol. 5, no. 1, Jun. 2023, doi: 10.47065/bits.v5i1.3597.

[14] L. M. Cendani and A. Wibowo, "Perbandingan Metode Ensemble Learning pada Klasifikasi Penyakit Diabetes," *J. Masy. Inform.*, pp. 33–44, 2022.

[15] G. Kyriakides and K. G. Margaritis, *Hands-On Ensemble Learning with Python: Build Highly Optimized Ensemble Machine Learning Models Using Scikit-learn and Keras*. 2019.

[16] A. S. Nugroho, A. B. Witarto, and D. Handoko, "Support Vector Machine-Teori dan Aplikasinya dalam Bioinformatika 1," *Proceeding Indones. Sci. Meeiting Cent. Japan.*, 2003, [Online]. Available: http://asnugroho.net

[17] E. Hasibuan and E. A. Heriyanto, "Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier," *J. Tek. dan Sci.*, pp. 13–24, 2022.

[18] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning," *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 1, Apr. 2021, doi: 10.28932/jutisi.v7i1.3216.