

PENGARUH KLASIFIKASI SENTIMEN PADA ULASAN PRODUK AMAZON BERBASIS REKAYASA FITUR DAN *K-NEAREST NEIGHBOR*

Nitami Lestari Putri^{*1}, Budi Warsito², Bayu Surarso³

¹²³Universitas Diponegoro, Semarang, Indonesia
Email: ¹nitamiputri38@gmail.com, ²budiwrst2@gmail.com, ³bayusurarso@yahoo.com
^{*}Penulis Korespondensi

(Naskah masuk: 14 Juni 2023, diterima untuk diterbitkan: 21 November 2023)

Abstrak

Ulasan *online* menjadi faktor penting yang mendorong konsumen untuk membeli barang di *e-commerce*. Dalam *e-commerce*, ulasan pelanggan sebelumnya dapat membantu pembeli membuat keputusan yang lebih baik dengan memberikan informasi tentang kualitas produk, kekuatan dan kelemahan, perilaku penjual, harga, dan waktu pengiriman. Namun, keberadaan ulasan palsu menimbulkan tantangan dalam menilai sentimen yang diungkapkan oleh pelanggan asli secara benar. Dalam penelitian ini, berfokus pada analisis sentimen dan bertujuan untuk mengeksplorasi peran sentimen dalam ulasan produk Amazon. Penelitian ini menggunakan kombinasi fitur dari konten ulasan dengan menerapkan klasifikasi *K-Nearest Neighbor* untuk mengklasifikasikan polaritas sentimen ulasan secara akurat. Dalam mengekstrak skor polaritas dari ulasan, penelitian ini menggunakan pendekatan analisis sentimen berbasis leksikon yaitu *Textblob Library* dan menetapkan label sentimen dari ulasan produk. Hasil dari pemodelan yang diusulkan mencapai tingkat akurasi sebesar 83% yang menunjukkan keefektifan pemodelan yang diusulkan dalam analisis sentimen. Hasil dari penelitian ini dapat membantu konsumen dalam membuat keputusan pembelian dan membantu penjual dalam meningkatkan nilai produk dan layanan mereka berdasarkan *feedback* yang diberikan oleh pelanggan.

Kata kunci: Ulasan produk Amazon, analisis sentimen, *Textblob Library*, kombinasi fitur, *K-Nearest Neighbor*

THE EFFECT OF SENTIMENT CLASSIFICATION ON AMAZON PRODUCT REVIEWS BASED ON FEATURE ENGINEERING AND *K-NEAREST NEIGHBOR*

Abstract

Online reviews are a significant influence encouraging consumers to buy things in e-commerce. In e-commerce, previous customer reviews can help buyers make better decisions by providing information about product quality, strengths and weaknesses, seller behaviour, prices, and delivery times. However, the existence of fake reviews poses a challenge in accurately assessing the sentiments of real customers. In this study, we focus on sentiment analysis and aim to explore the role of sentiment in Amazon product reviews. This study uses a combination of features from content reviews by correctly classifying review sentiment polarity using the K-Nearest Neighbor classification. In exploring polarity scores from reviews, this study uses a lexicon-based sentiment analysis approach, namely the Textblob Library that attaches sentiment labels from product reviews. The proposed modeling is effective in sentiment analysis as evidenced by the result, which have an accuracy rate of 83%. The results of this research can assist consumers in making purchasing decisions and assist sellers in increasing the value of their products and services based on feedback provided by customers.

Keywords: Amazon's product reviews, sentiment analysis, *Textblob Library*, Features Combination, *K-Nearest Neighbor*

1. PENDAHULUAN

Ulasan produk *online* memainkan peran utama dalam keberhasilan atau kegagalan bisnis *e-commerce*. Dalam sebuah transaksi, pembeli biasanya akan mencari tau informasi penggunaan produk atau jasa tersebut dari ulasan *online* yang diposting oleh pelanggan sebelumnya untuk

mendapatkan rekomendasi detail produk dan membuat keputusan pembelian (Alsubari et al., 2021). Ulasan *online* juga bermanfaat bagi calon pembeli dan vendor. Dengan ulasan yang diberikan oleh pembeli, vendor mampu merancang strategi pemasaran tambahan mereka. (Daiv et al., 2020). Salah satu *e-commerce* yaitu Amazon yang didirikan oleh Jeffrey P. Bezos pada tahun 1994 adalah *e-*

commerce paling terkenal di dunia saat ini. Nama Amazon berasal dari sungai Amazon, yang merupakan sungai terbesar di dunia (Wells, Danskin, & Ellsworth, 2018). Amazon telah memperluas bisnisnya dengan mencakup barang-barang ritel lainnya seperti kosmetik dan peralatan elektronik. Karena keberhasilannya, Amazon berhasil mendapatkan peringkat yang tinggi diantara toko ritel *online* lainnya yang terkenal di dunia (Tangmanee & Jongtavornvitaya, 2022). Platform *e-commerce* menawarkan beberapa keunggulan seperti proses pembelian yang lebih cepat, respon yang lebih cepat terhadap permintaan pembeli atau pasar, dan menyediakan beberapa mode pembayaran (Rodrigues et al., 2022).

Banyak ulasan yang dibuat oleh pengguna yang seringkali menyertakan opini sentimen yang kuat. Data ulasan ini sangat menjanjikan dan dapat digunakan baik oleh pelanggan maupun Perusahaan. Pelanggan dapat membaca ulasan untuk mengetahui lebih banyak tentang kualitas suatu produk. Namun, karena banyaknya ulasan, sulit untuk melihat dan membaca semua evaluasi konsumen secara pribadi untuk mendapatkan informasi yang berguna. Masalah ini membutuhkan teknik untuk dapat meninjau dan menganalisis data dengan mengekspresikan data dengan cara yang dapat dipahami oleh komputer (Sangeetha & Kumaran, 2023). Masalah lainnya yaitu terdapat ulasan palsu yang dapat menyebabkan kerugian finansial bagi pelaku bisnis *e-commerce* dan dapat menyesatkan konsumen untuk mengambil keputusan yang salah (Alsubari et al., 2021). Keberadaan ulasan palsu juga dapat mempersulit dalam mengetahui sentimen sebenarnya pada ulasan yang dibagikan oleh pembeli sebenarnya. Berdasarkan permasalahan yang ada, dibutuhkan analisis terhadap teks ulasan yang diberikan oleh pengguna *e-commerce* menggunakan teknik penambangan teks. Salah satu teknik penambangan teks yang populer yaitu teknik analisis sentimen. Dalam penjualan produk *online*, tujuan dari analisis sentimen yaitu untuk memahami perasaan seseorang tentang produk atau merek tertentu (Birim et al., 2022).

Dalam analisis sentimen, ada dua pendekatan yang dapat digunakan dalam mengekstrak sentimen dari ulasan *online*. Pertama, yaitu pendekatan secara *machine learning* yang secara otomatis mengklasifikasikan ulasan yang membutuhkan data pelatihan (Bonta, Kumaresh & Janardhan, 2019). Dalam penelitian yang dilakukan oleh Akter, dkk., algoritma *K-Nearest Neighbor* (KNN) digunakan untuk menganalisis sentimen ulasan produk "Daraz", sebuah situs web *e-commerce* yang ada di Bangladesh. Hasil penelitian ini menunjukkan penerapan dari algoritma KNN dengan vektorisasi TF-IDF menghasilkan akurasi sebesar 96,25% (Akter, Begum & Mustafa, 2021). Kedua, yaitu pendekatan berbasis leksikon yang membutuhkan kamus yang berisi leksikon yang telah ditentukan

sebelumnya dan informasi tentang polaritas kata-kata yang berhubungan dengan sentimen (Kim & Lim, 2021). Salah satu teknik analisis sentimen berbasis leksikon yang dapat digunakan yaitu *Textblob library*. Seperti penelitian yang dilakukan Qorib, dkk., yang menggunakan *Textblob library* untuk menghitung nilai sentimen yang kemudian melakukan vektorisasi kata menggunakan ekstraksi fitur TF-IDF dan mengklasifikasikannya menggunakan *LinearSVC* pada kumpulan data Twitter vaksinasi COVID-19 dan mendapatkan hasil akurasi yaitu 96,7% untuk model yang diusulkan (Qorib et al., 2023).

Pada penelitian sebelumnya, penentuan fitur dapat menjadi pertimbangan untuk dapat melakukan analisis ulasan produk. Daiv, dkk., pada penelitiannya menggabungkan beberapa fitur yang termasuk dalam fitur *review centric* dan *reviewer centric*. Fitur-fitur tersebut digunakan untuk mengidentifikasi ulasan palsu pada dataset ulasan produk Amazon. Kemudian, algoritma *Logistic Regression* digunakan untuk memodelkan probabilitas kelas yang ada. Hasil akurasi yang didapatkan pada penelitian ini sebesar 82,1% (Daiv et al., 2020). Penelitian lain yang dilakukan Birim, dkk., yaitu melakukan kombinasi fitur berdasarkan fitur *review centric* dan *reviewer centric* untuk pendeteksian ulasan palsu pada dataset ulasan produk Amazon. Penelitian ini juga menambahkan kombinasi fitur dengan menerapkan distribusi topik yang dihasilkan dari penerapan algoritma *Latent Dirichlet Allocation* (LDA). Kombinasi fitur yang diusulkan dievaluasi menggunakan klasifikasi *Random Forest*. Hasil akurasi yang didapatkan pada penelitian ini sebesar 81,57% (Birim et al., 2022).

Pada penelitian ini melakukan analisis sentimen pada dataset ulasan produk Amazon dengan mengkombinasikan fitur yang termasuk dalam fitur *review centric* dan *reviewer centric*. Kemudian, *Textblob library* digunakan untuk mendapatkan skor polaritas dari ulasan produk yang dimana skor tersebut dijadikan penentuan untuk label sentimen. Selanjutnya, hasil dari kombinasi fitur dievaluasi menggunakan algoritma *K-Nearest Neighbor* untuk menentukan model yang paling akurat. Tujuan penelitian ini untuk menentukan bagaimana sentimen memainkan peran dalam ulasan produk Amazon dan membantu penjual dan konsumen membuat keputusan yang tepat.

2. LANDASAN TEORI

4.1. *Textblob Library*

Textblob adalah teknik analisis sentimen berbasis leksikon yang populer dengan model pustaka yang tersedia dalam Python yang menyediakan pemrosesan teks yang disederhanakan. *Application Programming Interface* (API) yang sederhanakan dari *Textblob* memfasilitasi banyak pemrosesan teks umum dan tugas dari NLP seperti terjemahan bahasa, *Parts of Speech* (POS) tagging, tokenisasi, ekstraksi

frasa, klasifikasi, analisis sentimen, dan lain-lain (Aljedaani et al., 2022).

Pada *Textblob*, untuk menghitung nilai polaritas menggunakan persamaan (1), dimana skor polaritas +1 menunjukkan kelas positif dan -1 menunjukkan kelas negatif. *Textblob* menerapkan konsep rata-rata aritmatika sederhana untuk menghitung polaritas (\bar{P}) yang di mana jumlah polaritas dari kata ($\sum_{i=1}^n X_i$) dibagi menjadi beberapa kali nilai (n) yang merupakan kata yang muncul dalam kumpulan data leksikal (Junior et al., 2021).

$$\bar{P}(x) = \frac{1}{n} \sum_{i=1}^n X_i \quad (1)$$

4.2. CountVectorizer

CountVectorizer (CV) biasa digunakan untuk perhitungan numerik fitur kelas dan metode ekstraksi fitur teks (Yang et al., 2020). Vektor yang dihasilkan dari CV disejajarkan karena panjang vektor adalah panjang dari kamus kata (Zheng et al, 2020). Tabel 1 merupakan contoh dari matriks CV yang menghasilkan matriks dengan ukuran 3x5 yang berisikan kalimat seperti "book is good", "book is average", dan "book is nice". Berdasarkan kalimat tersebut, terdapat 3 dokumen dan 5 fitur berbeda seperti *book*, *is*, *good*, *average*, dan *nice*. Pada matriks, setiap nilai 1 dalam satu baris sesuai dengan keberadaan fitur dan 0 mewakili tidak adanya fitur dari dokumen tertentu.

Tabel 1. Skema matriks *CountVectorizer*

	Fitur1	Fitur2	Fitur3	Fitur4	Fitur5
Dokumen1	1	1	1	0	0
Dokumen2	1	1	0	1	0
Dokumen3	1	1	0	0	1

4.3. TF-IDF

Menurut Kaur, dkk., *Term Frequency – Inverse Document Frequency* (TF-IDF) terdiri dari dua metrik, *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). Tingkat kepentingan *term* terhadap dokumen dalam dataset dapat diukur dengan TF-IDF. Persamaan (2) menunjukkan TF-IDF, dimana frekuensi istilah diwakili dengan tf dan dihitung dari *count* (c), *term* (t) pada dokumen (d) dan direpresentasikan sebagai $(tf(t, d) = C_{td})$. Frekuensi kemunculan kata menjadi fitur biner yang dikonversi dengan menggunakan 1 yang berarti kata ada dalam dokumen dan 0 yang berarti kata tidak ada dalam dokumen.

$$tfidf = tf(t, d) \times idf(t, d) \quad (2)$$

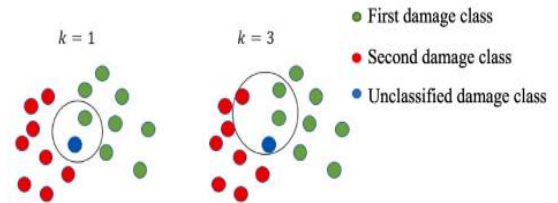
Inverse Document Frequency (idf) untuk kata w pada dokumen teks (t) dapat dilihat pada persamaan (3). T mewakili jumlah total dokumen korpus dan

$df(t)$ mewakili hitungan jumlah dokumen pada istilah (t) (Kaur et al., 2020).

$$idf(t, d) = 1 + \log \frac{T}{(1+df(t))} \quad (3)$$

4.4. K-Nearest Neighbor

Menurut Kiani, dkk., Algoritma *K-Nearest Neighbor* (KNN) adalah algoritma klasifikasi yang tidak membuat asumsi apapun pada distribusi data yang mendasari dan bekerja berdasarkan kesamaan fitur. Dalam KNN, titik data diklasifikasikan berdasarkan suara mayoritas tetangganya dengan titik data dialokasikan ke kelas yang paling umum diantara K tetangga terdekatnya. Selain detail teknik, KNN juga menghitung batas keputusan untuk lebih dari dua kelas untuk mengklasifikasikan poin baru. Jumlah tetangga K adalah *hyperparameter* yang menyesuaikan kompleksitas pengklasifikasi ini (Kiani, Camp & Pezeshk, 2019). Pada Gambar 2 menunjukkan secara sederhana ilustrasi cara kerja KNN. KNN dapat menetapkan kelas kerusakan ke kelas yang tidak diklasifikasikan (biru) berdasarkan suara mayoritas dari K tetangga terdekat di set pelatihan (kelas kerusakan hijau dan merah). Pada Gambar 1 ditunjukkan dengan menetapkan nilai $k=1$ dan $k=3$ (Ghiasi, Ng & Sheikh, 2022).



Gambar 1. Skema klasifikasi KNN 2 dimensi

Cosine similarity merupakan metrik yang populer dalam analisis teks. Ketika teks diproses sebagai sekumpulan kata, maka sekumpulan kata tersebut merupakan fitur dan *cosine similarity* memiliki keuntungan dengan tidak bergantung pada besarnya vektor fitur (Cunningham & Delany, 2021). *Cosine similarity* diwakili oleh persamaan (4) dimana θ adalah sudut vektor A dan B. A dan B mewakili vektor fitur, A_i^2 mewakili panjang vektor A dan B_i^2 mewakili panjang vektor B (Han et al., 2019).

$$Similarity = \text{Cos}(\theta) = \frac{\sum_{i=1}^p A_i \times B_i}{\sqrt{\sum_{i=1}^p A_i^2} \times \sqrt{\sum_{i=1}^p B_i^2}} \quad (4)$$

4.5. Confusion Matrix

Dalam *machine learning*, *confusion matrix* adalah alat evaluasi visual yang terdiri dari kolom dan baris yang menunjukkan hasil kelas yang sebenarnya dan prediksi (Xu, Zhang & Miao, 2020). Elemen

matriks dikarakterisasi berdasarkan label prediksi (positif dan negatif) dan hasil perbandingan prediksi dengan label kelas sebenarnya (*true* dan *false*) yaitu *true positives* (TP), *true negatives* (TN), *false positives* (FP), dan *false negatives* (FN). *Confusion matrix* dapat diterapkan pada sekumpulan metrik kinerja dari masalah klasifikasi untuk menilai suatu algoritma atau membandingkan performa algoritma yang berbeda. Berikut merupakan *confusion matrix* yang ditunjukkan pada Gambar 2.

		Predicted Class	
		Positive	Negative
Actual Class	Positive	TP	FN
	Negative	FP	TN

Gambar 2. *Confusion matrix*

Beberapa ukuran kinerja yang sering digunakan untuk mengevaluasi algoritma klasifikasi yang diterapkan pada *confusion matrix* yaitu akurasi, presisi, *recall*, dan *F1-score*. Berikut merupakan persamaan dari keempat metrik kinerja yang masing-masing disajikan pada persamaan (5), (6), (7), dan (8) (Alsubari et al., 2021).

$$\text{Akurasi} = \frac{TP+TN}{FP+FN+TP+TN} \times 100 \quad (5)$$

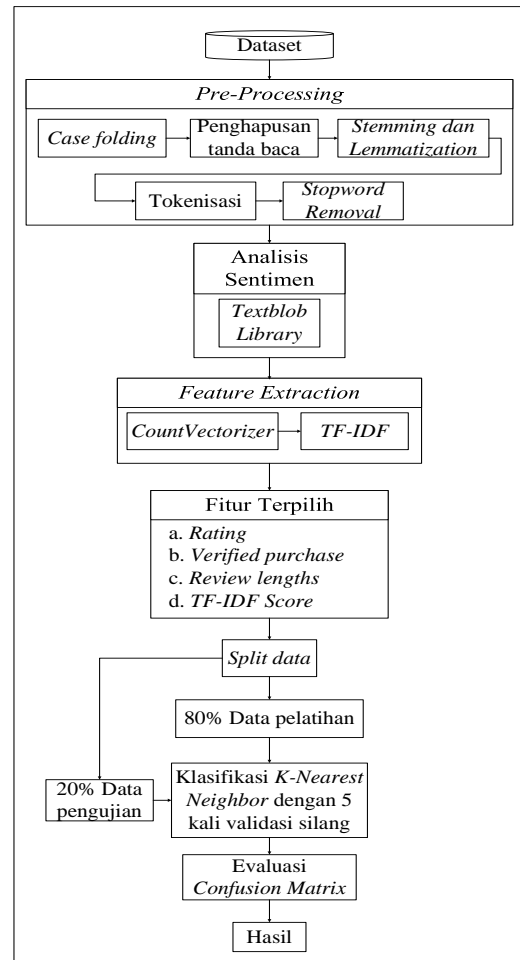
$$\text{Presisi} = \frac{TP}{TP+FP} \times 100 \quad (6)$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100 \quad (7)$$

$$\text{F1-Score} = 2 * \frac{\text{precision} \times \text{sensitivity}}{\text{precision} + \text{sensitivity}} \times 100 \quad (8)$$

3. METODE PENELITIAN

Penelitian ini menggunakan perangkat keras *Processor Intel Core i7* dan perangkat lunak *Windows 10 x64* sebagai sistem operasi. Bahasa pemrograman yang digunakan yaitu *Phyton 3.9.2 x64* dan *Jupiter Notebook*. Penelitian ini menggunakan metodologi yang didasarkan pada alur proses penelitian yang digambarkan pada Gambar 3. Alur proses penelitian terdiri dari beberapa tahapan yang masing-masing dijelaskan sebagai berikut:



Gambar 3. Alur Proses Penelitian

4.1. Pengumpulan Data

Dataset ulasan produk berbahasa Inggris standar Amazon digunakan untuk penelitian ini. Dataset ini terdiri dari 21.000 ulasan. Setiap ulasan memiliki fitur-fitur seperti ID *document*, label yang menunjukkan ulasan palsu dan asli, *rating*, *verified purchase*, *product category*, ID *product*, *product title*, *review title*, dan *review text*. Dataset ini diperoleh dari repositori GitHub Aayush Saxena pada tautan berikut <https://github.com/aayush210789/Deception-Detection-on-Amazon-reviews-dataset> (Birim et al., 2022).

4.2. Pre-Processing

Data yang sudah diperoleh kemudian dilakukan *pre-processing* yang bertujuan untuk merekonstruksi teks menjadi bentuk yang lebih mudah dicerna untuk algoritma pembelajaran mesin. Terdapat 5 tahapan dalam *pre-processing* pada penelitian ini. Tahapan-tahapan tersebut antara lain *case folding*, penghapusan tanda baca, *stemming* dan *lemmatization*, tokenisasi dengan menggunakan unigram, dan *stopword removal*.

4.3. Analisis Sentimen

Analisis sentimen pada penelitian ini menggunakan pendekatan berbasis leksikon. Salah satu teknik analisis sentimen berbasis leksikon yang dapat digunakan yaitu *Textblob library*. *Textblob library* digunakan pada penelitian ini untuk mencari nilai polaritas dari setiap kata dalam ulasan teks. Setelah memperoleh nilai polaritas, kemudian ditentukan label sentimen positif, negatif, dan netral.

4.4. Feature Extraction

Penelitian ini menggunakan *CountVectorizer* (CV) untuk menghitung frekuensi kata yang muncul dan TF-IDF untuk menghitung nilai bobot setiap kata dalam ulasan. Pada proses CV, dipilih 1400 kata yang paling sering muncul dalam ulasan untuk membentuk vektor. Setelah data teks divektorisasi dan dihitung jumlah frekuensi kemunculan katanya, selanjutnya menghitung nilai bobot setiap kata menggunakan TF-IDF.

4.5. Pemilihan Fitur

Pada penelitian sebelumnya telah banyak diusulkan berbagai pendekatan untuk mengidentifikasi ulasan berdasarkan fitur-fitur *review centric* dan *reviewer centric*. Fitur *review centric* menggunakan fitur yang diekstraksi berdasarkan setiap kata dari ulasan dan fitur *reviewer centric* bertujuan untuk menemukan karakteristik perilaku pengulas dari ulasan yang mereka tulis (Birim et al., 2022). Fitur-fitur tersebut dapat digunakan untuk mengidentifikasi berbagai persoalan untuk analisis teks. Pada penelitian sebelumnya, fitur *review centric* dan *reviewer centric* digunakan untuk mengidentifikasi ulasan palsu pada dataset Amazon. Pada penelitian ini, fitur *review centric* dan *reviewer centric* digunakan untuk mengidentifikasi sentimen pada ulasan produk. Berikut merupakan fitur-fitur terpilih yang digolongkan sebagai fitur *review centric* dan *reviewer centric* untuk mengidentifikasi sentimen pada dataset ulasan produk Amazon yang dijelaskan sebagai berikut:

a. Review Centric

Fitur *review centric* atau fitur yang berpusat pada ulasan bertujuan untuk menangkap informasi tekstual dan karakteristik dari setiap ulasan. Pada penelitian ini, hasil dari ekstraksi fitur menggunakan *CountVectorizer* dan TF-IDF digolongkan sebagai fitur *review centric*. Hal ini karena hasil dari ekstraksi fitur menunjukkan frekuensi atau pentingnya suatu kata dalam ulasan dan dapat menangkap informasi tekstual dan karakteristik dari setiap ulasan.

b. Reviewer Centric

Fitur *reviewer centric* atau fitur yang berpusat pada pengulas berfokus pada mengidentifikasi perilaku dan karakteristik pengulas. Pada penelitian ini *rating*, *verified purchase*, dan *review lengths* digolongkan sebagai *reviewer centric*. Berikut merupakan deskripsi dari ketiga fitur menurut (Daiv et al., 2020) sebagai berikut:

1. Rating

Fitur *rating* memuat nilai peringkat produk yang diberikan oleh pengguna *e-commerce* yang terdiri dari peringkat 1 sampai dengan 5. Fitur ini mewakili kepuasan atau ketidakpuasan pengguna terhadap produk yang mereka beli. Fitur ini dapat digunakan untuk memvalidasi bahwa ulasan yang ditulis dan penilaian yang diberikan oleh pengguna ditujukan hanya satu arah dan tidak bertentangan.

2. Verified Purchase

Fitur *verified purchase* pada dataset ulasan produk amazon merupakan fitur yang memverifikasi pengulas yang membeli produk di Amazon telah membeli produk tersebut dan tidak menerima produk dengan diskon besar. Fitur ini dapat digunakan untuk mengetahui pembeli mana yang benar-benar membeli produk dan menggunakannya serta mengetahui sentimen dari ulasan yang mereka tulis.

3. Review Lengths

Fitur *review lengths* atau panjang ulasan dapat digunakan untuk melatih model *machine learning*. Pada penelitian analisis sentimen, panjang ulasan dapat digunakan untuk mengetahui karakteristik pengulas dari panjang dan pendeknya ulasan yang ditulis.

Pada penelitian ini fitur-fitur yang termasuk *review centric* dan *reviewer centric* digunakan untuk mengidentifikasi sentimen pada ulasan produk Amazon. Fitur-fitur tersebut antara lain TF-IDF score, *rating*, *verified purchase*, dan *review lengths*. Fitur-fitur tersebut dipilih sebagai prediktor dan digunakan untuk melatih model *machine learning*. Pemilihan fitur-fitur tersebut diharapkan dapat meningkatkan kinerja algoritma klasifikasi yang digunakan dan meningkatkan nilai akurasi.

4.6. Klasifikasi K-Nearest Neighbor

Sebelum masuk ke tahap klasifikasi, data dibagi menjadi 2 bagian. Dari total 21.000 data, 80% atau 16.800 data merupakan data pelatihan dan 20% atau 4.200 data merupakan data pengujian. Setelah data dibagi, data pelatihan dan pengujian digunakan untuk melatih dan menguji algoritma *K-Nearest Neighbor* (KNN) untuk mencari parameter *k* dan akurasi optimal. Proses pelatihan dan pengujian algoritma KNN dilakukan dan dievaluasi menggunakan 5 kali validasi silang. Pada proses 5 kali validasi silang, setiap data dibagi menjadi 5 bagian yang masing-

masing bagian berjumlah 4.200 data yang dimana dari 5 bagian, 4 bagian diantaranya merupakan data pelatihan dan 1 bagian lainnya merupakan data pengujian. Validasi silang dilakukan sebanyak 5 kali iterasi sampai mendapatkan hasil prediksi terbaik dari algoritma KNN. Proses evaluasi 5 kali validasi silang ditunjukkan pada Gambar 4.

	A Percobaan 1 (4.200 data)	B Percobaan 2 (4.200 data)	C Percobaan 3 (4.200 data)	D Percobaan 4 (4.200 data)	E Percobaan 5 (4.200 data)
Percobaan 1	A (data pengujian)	BCDE (data pelatihan)			
Percobaan 2	A	B (data pengujian)	CDE (data pelatihan)		
Percobaan 3	AB (data pelatihan)		C (data pengujian)	DE (data pelatihan)	
Percobaan 4	ABC (data pelatihan)			D (data pengujian)	E (data pelatihan)
Percobaan 5	ABCD (data pelatihan)				E (data pengujian)

Gambar 4. Evaluasi 5 kali validasi silang

Pada penelitian ini, *cosine similarity* digunakan untuk menghitung kesamaan kosinus antara vektor fitur untuk menentukan kesamaan antar *instance*. Pengujian dilakukan dengan menetapkan nilai k dari k=1 sampai dengan k=13 untuk mengetahui nilai akurasi yang paling optimal.

4.7. Evaluasi Confusion Matrix

Penelitian ini menggunakan evaluasi *confusion matrix* untuk menghitung metrik kinerja. Empat metrik kinerja tersebut adalah akurasi, presisi, *recall*, dan *F1-score*.

4. HASIL DAN PEMBAHASAN

4.1. Pre-Processing

Beberapa tahapan pra-pengolahan teks yang diterapkan antara lain *case folding*, penghapusan tanda baca, *stemming* dan *lemmatization*, tokenisasi, dan *stopword removal*. Pada tahap *case folding*, setiap huruf diubah menjadi huruf kecil. Proses *case folding* digambarkan pada Tabel 2.

Tabel 2. Proses Case Folding

Sebelum Case Folding	Setelah Case Folding
I have had my camera for about 2 weeks now. The pictures are great, it is so easy to use. I love it!!	i have had my camera for about 2 weeks now. the pictures are great, it is so easy to use. i love it!!

Tahap selanjutnya yaitu penghapusan tanda baca. Berikut merupakan proses penghapusan tanda baca yang disajikan pada Tabel 3.

Tabel 3. Proses Penghapusan Tanda Baca

Sebelum penghapusan tanda baca	Setelah penghapusan tanda baca
i have had my camera for about 2 weeks now. the pictures are great, it is so easy to use. i love it!!	i have had my camera for about 2 weeks now the pictures are great it is so easy to use i love it

Pada penelitian yang dilakukan Qorib, dkk., yaitu menggabungkan *stemming* dan *lemmatization* untuk meningkatkan kinerja model yang diusulkan.

Penggabungan kedua teknik ini menghasilkan tingkat akurasi tertinggi 96,7% (Qorib et al., 2023). Berdasarkan penelitian sebelumnya, pada penelitian ini teknik normalisasi kosakata yaitu *stemming* dan *lemmatization* digabungkan untuk meningkatkan kinerja model yang diusulkan. Pada tahap *stemming*, kata diubah menjadi bentuk dasarnya. Proses *stemming* bertujuan untuk menghilangkan imbuhan prefiks, sufiks, dan konfiks yang ada di setiap kata. Kemudian dilanjutkan dengan tahap *lemmatization*. *Lemmatization* adalah proses mereduksi kata yang diinfleksikan dengan benar dan memastikan bahwa kata dasar tersebut memiliki bahasa yang sesuai kamus. *Lemmatization* didasarkan pada kosa kata dan bentuk kata aslinya (Qorib et al., 2023). Berikut merupakan proses *stemming* dan *lemmatization* yang disajikan pada Tabel 4.

Tabel 4. Proses stemming dan Lemmatization

Sebelum proses stemming dan lemmatization	Setelah proses stemming dan lemmatization
i have had my camera for about 2 weeks now the pictures are great it is so easy to use i love it	i have had my camera for about 2 week now the pictur are great it is so easi to use i love it

Tahap selanjutnya yaitu tokenisasi. Tokenisasi dilakukan dengan melihat setiap spasi dalam ulasan. Setiap kata dapat dipecah berdasarkan spasi tersebut (Birim et al., 2022). Penelitian ini menggunakan unigram untuk menghilangkan pembatas dan spasi kata. Hasil dari tokenisasi kemudian disimpan dalam bentuk *array* untuk dianalisis lebih lanjut (Ali et al., 2019). Berikut merupakan proses tokenisasi yang disajikan pada Tabel 5.

Tabel 5. Proses Tokenisasi

Sebelum proses tokenisasi	Setelah proses tokenisasi
i have had my camera for about 2 week now the pictur are great it is so easi to use i love it	['i', 'have', 'had', 'my', 'camera', 'for', 'about', '2', 'week', 'now', 'the', 'pictur', 'are', 'great', 'it', 'is', 'so', 'easi', 'to', 'use', 'i', 'love', 'it']

Selanjutnya adalah proses *stopword removal*. Berikut merupakan proses *stopword removal* yang disajikan pada Tabel 6.

Tabel 6. Proses Stopword Removal

Sebelum proses stopword removal	Setelah proses stopword removal
['i', 'have', 'had', 'my', 'camera', 'for', 'about', '2', 'week', 'now', 'the', 'pictur', 'are', 'great', 'it', 'is', 'so', 'easi', 'to', 'use', 'i', 'love', 'it']	['camera', '2', 'week', 'pictur', 'great', 'easi', 'use', 'love']

4.2. Analisis Sentimen Textblob

Tahap selanjutnya melakukan analisis sentimen menggunakan *Textblob library*. *Textblob library* digunakan untuk mencari nilai polaritas dari setiap ulasan untuk menangkap sentimen dari ulasan

tersebut. Setelah nilai polaritas didapatkan, selanjutnya ditetapkan label sentimen berdasarkan nilai polaritasnya. Penentuan ambang batas untuk label sentimen positif, negatif, dan netral ditetapkan apabila skor polaritas >0 maka dinyatakan positif, apabila skor polaritas <0 maka dinyatakan negatif, dan apabila skor polaritas $=0$ maka dinyatakan netral. Berikut merupakan perolehan nilai polaritas dan label sentimen pada ulasan produk Amazon yang disajikan pada Tabel 7.

Tabel 7. Proses *Stopword Removal*

Ulasan	Skor polaritas	Label sentimen
['camera', '2', 'week', 'pictur', 'great', 'easi', 'use', 'love']	0.65	Positif
['unfortun', 'didnt', 'work', 'made', 'sick', 'throw', 'onc', 'two', 'time', 'tri']	-0.71	Negatif
['wa', 'look', 'inexpen', 'desk', 'calcolatur', 'work', 'doe', 'everyth', 'need', 'onli', 'issu', 'tilt', 'slightli', 'one', 'side', 'hit', 'ani', 'key', 'rock', 'littl', 'bit', 'big', 'deal']	0	Netral

Pada Tabel 8 merupakan total keseluruhan jumlah sentimen. Dapat dilihat bahwa dari total 21.000 ulasan, 17.275 diantaranya merupakan ulasan positif, 2.154 merupakan ulasan negatif, dan 1.571 merupakan ulasan netral.

Tabel 8. Total jumlah label sentimen

Label sentimen	Jumlah
Positif	17.275
Negatif	2.154
Netral	1.571
Total:	21.000

4.3. Feature Extraction

Tahap ekstraksi fitur pada penelitian ini melibatkan *CountVectorizer* (CV) dan TF-IDF. Pada CV, dipilih 1400 kata teratas yang paling sering muncul di dalam ulasan untuk membentuk vektor fitur. Berikut merupakan 10 kata teratas dari 1400 kata yang paling sering muncul di dalam ulasan yang disajikan pada Tabel 9.

Tabel 9. 10 kata yang paling sering muncul di dalam ulasan

No	Kata	Jumlah kemunculan
1	<i>section</i>	24222
2	<i>sauc</i>	10697
3	<i>mattress</i>	8631
4	<i>resist</i>	7829
5	<i>everywher</i>	6488
6	<i>foot</i>	6293
7	<i>effect</i>	6037
8	<i>zoom</i>	5376
9	<i>run</i>	5186
10	<i>poorli</i>	4846

Berdasarkan Tabel 9, dari 1400 kata yang digunakan, kata "*section*" merupakan kata yang sering muncul dengan jumlah kemunculan 24222 kata. Kedua adalah kata "*sauc*" dengan jumlah kemunculan 10697 kata. Ketiga adalah kata "*mattress*" dengan jumlah kemunculan 8631 kata. Kemudian dilanjutkan dengan kata "*resist*", "*everywher*", "*foot*", "*effect*", "*zoom*", "*run*", dan "*poorly*" dengan jumlah kemunculan kata masing-masing adalah 7829 kata, 6488 kata, 6293 kata, 6037 kata, 5376 kata, 5186 kata, dan 4846 kata. Selanjutnya adalah menghitung bobot setiap kata menggunakan TF-IDF. Transformasi dilakukan pada *sparse matrix* yang sudah dibentuk sebelumnya dari data teks berupa vektor menggunakan CV. Tabel 10 menunjukkan implementasi TF-IDF pada 5 kata yang dipilih.

Tabel 10. Hasil penerapan TF-IDF

Dokumen ke-	Indeks kata	Bobot kata
0	1237	0.294
0	1233	0.115
0	1127	0.319
0	1030	0.385
0	938	0.229

4.4. Klasifikasi *K-Nearest Neighbor*

Sebelum masuk ke tahap klasifikasi, data dibagi menjadi 80% data pelatihan dan 20% data pengujian. Pada tahap sebelumnya, dipilih 4 fitur yang tergolong sebagai fitur *review centric* dan *reviewer centric*. Fitur-fitur tersebut antara lain TF-IDF *score*, *rating*, *verified purchase*, dan *review lengths*. Fitur-fitur tersebut digunakan sebagai prediktor dan juga untuk melatih algoritma klasifikasi *K-Nearest Neighbor* (KNN).

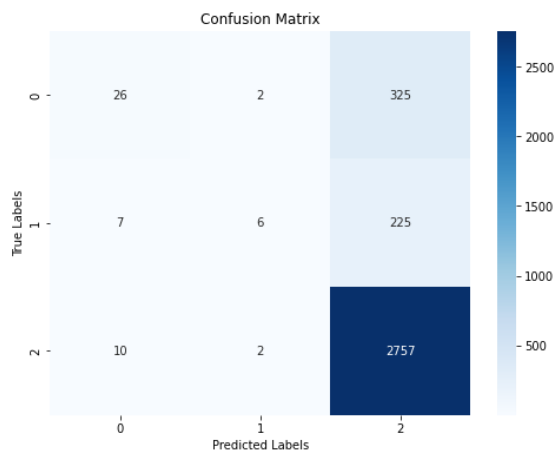
Klasifikasi dilakukan dengan menggunakan lima kali validasi silang untuk dapat mengevaluasi kinerja model. Kemudian, Pengujian dilakukan dengan menetapkan nilai $k=1$ sampai $k=13$ untuk KNN. Hasil klasifikasi disajikan dalam bentuk *confusion matrix* yang menetapkan empat metrik kinerja yaitu akurasi, presisi, *recall*, dan F1-score. Hasil dari penerapan KNN ditunjukkan pada Tabel 11.

Tabel 11. Hasil kinerja model

K=	Akurasi	Presisi	Recall	F1-Score
1	76.2%	73.3%	76.2%	74.6%
2	71.0%	74.2%	71.0%	72.4%
3	80.8%	74.2%	80.8%	76.3%
4	80.4%	74.1%	80.4%	76.3%
5	82.1%	75.4%	82.1%	76.5%
6	82.2%	76.3%	82.2%	76.8%
7	82.6%	76.6%	82.6%	76.3%
8	82.7%	77.0%	82.7%	76.4%
9	82.8%	77.6%	82.8%	76.1%
10	83.0%	79.3%	83.0%	76.5%
11	82.7%	76.7%	82.7%	75.9%
12	82.7%	76.5%	82.7%	76.0%
13	82.6%	77.0%	82.6%	75.7%

Pada Tabel 11 dapat dilihat model yang diusulkan memperoleh nilai terbaik pada $k=10$

dengan nilai akurasi sebesar 83,0%, nilai presisi 79.3%, nilai *recall* 83.0%, dan pada $k=6$ dengan nilai *F1-score* 76.8%. Berikut merupakan *confusion matrix* untuk nilai $k=10$ yang dapat dilihat pada Gambar 5.



Gambar 5. Hasil *confusion matrix*

Pada Gambar 5, 0 adalah representasi label negatif, 1 merupakan label netral, dan 2 merupakan label positif. Pada "True Labels" merupakan label sebenarnya dan "Predicted Labels" merupakan sentimen yang diprediksi dari ulasan produk Amazon. *Confusion matrix* menunjukkan bahwa $26+6+2.757=2.789$ ulasan diprediksi dengan sentimen yang sama dengan label aslinya. Kemudian, $2+325+225+7+10+2=571$ ulasan diprediksi dengan sentimen yang salah. Hasil kinerja model yang diusulkan kemudian dibandingkan dengan penelitian terdahulu yang menggunakan dataset ulasan produk standar Amazon. Terlepas dari perbedaan dengan penelitian sebelumnya, tujuan penelitian ini adalah untuk meningkatkan pemahaman tentang pengaruh sentimen pada ulasan produk Amazon. Perbandingan dilakukan untuk mengamati kontribusi yang berbeda dengan penelitian yang diusulkan untuk menunjukkan kemampuan rekayasa fitur dan algoritma KNN dalam analisis sentimen. Berikut merupakan perbandingan hasil akurasi dengan penelitian sebelumnya yang dapat dilihat pada Tabel 12.

Tabel 12. Perbandingan hasil dengan penelitian sebelumnya

Penelitian dari:	Fitur yang digunakan	Metode yang digunakan	Akurasi
Daiv, dkk (2020)	Fitur ulasan konten, <i>rating</i> , <i>verified purchase</i> , dan <i>review lengths</i>	<i>Logistic Regression</i>	82.1%
Birim, dkk (2022)	<i>Topic distribution</i> , <i>verified purchase</i> , dan <i>review lengths</i>	<i>Random Forest</i>	81.5%
Penelitian yang diusulkan	<i>TF-IDF score</i> , <i>rating</i> , <i>verified purchase</i> , <i>review lengths</i>	<i>K-Nearest Neighbor</i>	83.0%

Pada Tabel 12, fitur yang digunakan pada penelitian ini untuk analisis sentimen menunjukkan akurasi yang cukup baik dibandingkan dengan penelitian terdahulu. Akurasi yang didapatkan pada penelitian ini sebesar 83,0%. Hal ini membuktikan bahwa kombinasi fitur yang digunakan efektif dalam mengklasifikasikan polaritas sentimen ulasan secara akurat. Akurasi yang didapatkan dalam penelitian ini menunjukkan bahwa kombinasi fitur yang digunakan dapat menangkap aspek penting dari sentimen pelanggan yang diungkapkan dalam ulasan produk Amazon. Penerapan KNN sebagai pengklasifikasi mampu mengklasifikasikan sentimen ulasan secara akurat dengan memanfaatkan kemampuan pengenalan pola dan klasifikasi berbasis kedekatan.

5. KESIMPULAN

Penelitian ini mengevaluasi peran analisis sentimen dalam ulasan produk Amazon dengan memanfaatkan kombinasi fitur dan penerapan klasifikasi KNN. Kombinasi fitur yang dipilih antara lain fitur *TF-IDF score*, *rating*, *verified purchase*, dan *review lengths*. Penelitian ini bertujuan untuk mengidentifikasi sentimen yang dinyatakan dalam ulasan secara akurat. Hasil menunjukkan bahwa kombinasi fitur yang dipilih dengan penerapan klasifikasi KNN memberikan tingkat akurasi yang signifikan sebesar 83,0%. Hal ini menunjukkan keefektifan pendekatan yang diusulkan dalam mengklasifikasikan sentimen ulasan secara akurat.

Dari hasil akurasi yang sudah diperoleh, maka model yang diusulkan efektif untuk mengidentifikasi ulasan produk Amazon. Hasil ini menunjukkan potensi untuk memanfaatkan fitur yang dipilih dan klasifikasi KNN dalam tugas analisis sentimen dan dapat membantu konsumen dan penjual dalam memahami sentimen pelanggan. Dengan memahami sentimen yang diungkapkan oleh pelanggan, baik konsumen maupun penjual dapat membuat keputusan yang tepat mengenai pembelian dan meningkatkan nilai produk yang dijual di *e-commerce*.

DAFTAR PUSTAKA

- AKTER, M.T., BEGUM, M., & MUSTAFA, R., 2021. Bengali Sentiment Analysis of E-Commerce Product Reviews Using K-Nearest Neighbors. 2021 International Conference on Information And Communication Technology For Sustainable Development (ICICT4SD), [online] pp. 40-44. Available at : <<https://ieeexplore.ieee.org/document/9396910>>.
- Ali, F., KWAK, D., KHAN, P., EL-SAPPAGH, S., ALI, A., ULLAH, S., KIM, K, H., KWAK,

- K, S., 2019. Transportation Sentiment Analysis Using Word Embedding And Ontology-Based Topic Modelling. *Knowledge-Based Systems*, [online] 174, pp. 27-44. Available at: <<https://doi.org/10.1016/j.knosys.2019.02.033>>.
- ALJEEDANI, W., RUSTAM, F., MKAOUER, M.W., GHALLAB, A., RUPAPARA, V., WASHINGTON, P.B., LEE, E., & ASHRAF, I., 2022. Sentiment Analysis on Twitter Data Integrating Textblob And Deep Learning Models: The Case of US Airline Industry. *Knowledge-Based Systems*, [online] 255, p.109780. Available at: <<https://doi.org/10.1016/j.knosys.2022.109780>>.
- ALSUBARI, S.N., DESMUKH, S.N., ALADHAILEH, M.H., ALSAADE, F.W., & ALDHYANI, T.H.H., 2021. Development of Integrated Neural Network Model For Identification of Fake Reviews in E-Commerce Using Multidomain Datasets. *Applied Bionics And Biomechanics*, [online] 2021, pp. 1-11. Available at : <<https://doi.org/10.1155/2021/5522574>>.
- BIRIM, S.O., KAZANCOGLU, I., MANGLA, S.K., KAHRAMAN, A., KUMAR, S., & KAZANCOGLU, Y., 2022. Detecting Fake Review Through Topic Modelling. *Journal of Bussiness Research*, [online] 149, pp. 884-900. Available at : <<https://doi.org/10.1016/j.jbusres.2022.05.081>>.
- BONTA, V., KUMARESH, N., & JANARDHAN, N., 2019. A Comprehensive Study on Lexicon Based Approaches For Sentiment Analysis. *Asian Journal of Computer Science And Technology*, [online] 8, pp. 1-6. Available at : <<https://www.researchgate.net/publication/333602124>>.
- CUNNINGHAM, P., dan DELANY, S.J., 2021. K-Nearest Neighbour Classifiers-A Tutorial. *ACM Computing Surveys (CSUR)*, [online] 54(6), pp. 1-25. Available at: <<https://dl.acm.org/doi/10.1145/3459665>>.
- DAIV, K., LACHAKE, M., JAGTAP, P., DHARIWAL, S., & GUTTE, P.V., 2020. An Approach To Detect Fake Reviews Based On Logistic Regression Using Review-Centric Features. *International Research Journal of Engineering And Technology (IRJET)*, 07, pp. 2107-2112.
- GHIASI, A., NG, C.T., & SHEIKH.A.H., 2022. Damage Detection of In-Service Steel Railway Bridges Using a Fine K-Nearest Neighbor Machine Learning Classifier. *Structures*, [online] 45, pp. 1920-1935. Available at: <<https://doi.org/10.1016/j.istruc.2022.10.019>>.
- HAN, X., YANG, G., QU, S., ZHANG, G., CHI, M., 2019. A Weighted Algorithm Based on Physical Distance And Cosine Similarity For Indoor Localization. 2019 14th IEEE Conference on Industrial Electronics And Applications (ICIEA), [online] pp. 179-183. Available at: <<https://ieeexplore.ieee.org/abstract/document/8833982>>.
- JUNIOR, G.A.D.O., JR, R.T.D.S., ALBUQUERQUER, R.D.O., & VILLALBA, L.J.G., 2021. Adversarial Attack on a Lexical Sentiment Analysis Classifier. *Computer Communications*, [online] 174, pp. 154-171. Available at: <<https://doi.org/10.1016/j.comcom.2021.04.026>>.
- KAUR, S., KUMAR, P., & KUMARAGURU, P., 2020. Automating Fake News Detection System Using Multi-Level Voting Model. *Soft Computing*, 24, pp. 9049-9069.
- KIANI, J., CAMP, C., & PEZESHK, S., 2019. On The Application of Machine Learning Techniques To Derive Seismic Fragility Curves. *Computers And Structures*, [online] 218, pp. 108-122. Available at: <<https://doi.org/10.1016/j.compstruc.2019.03.004>>.
- KIM, J., dan LIM, C., 2021. Customer Complaints Monitoring With Customer Review Data Analytics: An Integrated Method of Sentiment And Statistical Process Control Analysis. *Advanced Engineering Informatics*, [online] 49, p. 101304. Available at : <<https://doi.org/10.1016/j.aei.2021.101304>>.
- QORIB, M., OLADUNNI, T., DENIS, M., OSOSANYA, E., & COTAE, P., 2022. COVID-19 Vaccine Hesitancy: Text Mining, Sentiment Analysis And Machine Learning on COVID-19 Vaccination Twitter Dataset. *Expert Systems With Applications*. [online] 212, p. 118715. Available at: <<https://doi.org/10.1016/j.eswa.2022.118715>>.
- RODRIGUES, V.F., POLICARPO, L.M., SILVEIRA, D.E.D., RIGHI, R.D.R., COSTA, C.A.D., BARBOSA, J.L.V., & ANTUNES, R.S., 2022. Fraud Detection And Prevention in E-Commerce: A Systematic Literature Review. *Electronic Commerce Research And Applications*, [online] 56, p.101207. Available at : <<https://doi.org/10.1016/j.eelerap.2022.101207>>.

- SANGEETHA, J., dan KUMARAN, U., 2023. Sentiment Analysis of Amazon User Review Using a Hybrid Approach. *Measurement: Sensors*. [online] 27, p.100790. Available at: <<https://doi.org/10.1016/j.measen.2023.100790>>.
- TANGMANEE, C., dan JONGTAVORNVITAYA, C., 2022. Visitors At Amazon.com: A Two-Level Exploration Into Their Browsing And Spending Behaviors. *International Journal of E-Business And E-Government Studies*, 14, pp. 225-245.
- WELLS, J.R., DANSKIN, G., & ELLSWORTH, G., 2018. Amazon.com 2018. Harvard Business School. [online] Available at : <<https://www.academia.edu/38615615/Amazon>>.
- XU, J., ZHANG, Y., & MIAO, D., 2020. Three-Way Confusion Matrix For Classification: A Measure Driven View. *Information Sciences*, [online] 507, pp. 772-794. Available at: <<https://doi.org/10.1016/j.ins.2019.06.064>>.
- YANG, J.S., ZHAO, C.Y., YU, H.T., & CHEN, H.Y., 2020. Use GBDT To Predict The Stock Market. *Proceedings 2019 International Conference on Identification, Information And Knowledge in The Internet of Things (IIKI2019)*, [online] 174, pp. 161-171. Available at : <<https://doi.org/10.1016/j.procs.2020.06.071>>.
- ZHENG, W., GAO, J., WU, X., LIU, F., XUN, Y., LIU, G., & CHEN, X., 2020. The Impact Factors on The Performance of Machine Learning-Based Vulnerability Detection: A Comparative Study. *The Journal of Systems & Software*, [online] 168, p. 110659. Available at: <<https://doi.org/10.1016/j.jss.2020.110659>>.