



Available online at : <http://bit.ly/InfoTekJar>

InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan

ISSN (Print) 2540-7597 | ISSN (Online) 2540-7600



Analisis Sentimen Produk Permainan Menggunakan Metode TF-IDF Dan Algoritma K-Nearest Neighbor

Rifki Kosasih¹, Anggi Alberto²

¹ Pusat Studi Komputasi Matematika Universitas Gunadarma, Jl. Margonda Raya No. 100, Depok 16424, Indonesia

² Fakultas Teknologi Informasi Universitas Gunadarma, Jl. Margonda Raya No. 100, Depok 16424, Indonesia

KEYWORDS

Analisis Sentimen, TF-IDF, KNN, Case Folding, Tokenizing

CORRESPONDENCE

Phone: +62 85281834545

E-mail: rifki_kosasih@staff.gunadarma.ac.id

ABSTRACT

On online shopping sites, there is a comment or rating column from buyers who have made transactions on the product. With the product rating feature based on this assessment, the buyer can find out how good or bad the product. However, a problem arises where there are some buyers who give negative comments with a five-star rating, this causes the product rating feature based on ratings to be less good. Therefore, to determine the quality of the product, sentiment analysis was carried out using the TF-IDF and K-Nearest Neighbor (KNN) methods based on reviews from buyers. The data collected is 1000 reviews which are divided into 700 training data and 300 test data. The next stage is preprocessing text such as case folding (converting uppercase letters to lowercase), tokenizing (separating single word sentences), stopwords (removing tokenizing conjunctions that have nothing to do with sentiment analysis), stemming (changing words into basic word forms) and word weighting with TF-IDF. The last stage is to classify using the K Nearest Neighbor (K-NN) method. Based on the classification obtained an accuracy rate of 79.3333%.

ABSTRAK

Pada situs belanja online, terdapat kolom komentar atau rating dari pembeli yang telah melakukan transaksi pada produk tersebut. Dengan adanya fitur penilaian produk berdasarkan rating tersebut, pihak pembeli dapat mengetahui seberapa baik atau buruknya produk tersebut. Akan tetapi muncul permasalahan dimana ada beberapa pembeli memberikan komentar negatif dengan rating sebesar lima bintang ataupun sebaliknya, hal tersebut menyebabkan fitur penilaian produk berdasarkan rating menjadi kurang baik. Oleh karena itu untuk dapat mengetahui kualitas produk tersebut dilakukan analisis sentimen dengan metode TF-IDF dan K-Nearest Neighbor (KNN) berdasarkan ulasan dari pembeli. Data yang dikumpulkan adalah 1000 ulasan yang dibagi menjadi 700 data latih dan 300 data uji. Tahapan selanjutnya dilakukan teks preprocessing seperti case folding (mengubah huruf besar menjadi kecil), tokenizing (pemisahan kalimat menjadi kata tunggal), stopword (menghilangkan kata sambung hasil tokenizing yang tidak ada hubungannya dalam analisis sentimen), stemming (mengubah kata ke bentuk kata dasar) dan pembobotan kata dengan TF-IDF. Tahapan terakhir adalah melakukan klasifikasi dengan menggunakan metode K Nearest Neighbor (K-NN). Berdasarkan hasil klasifikasi diperoleh tingkat akurasi sebesar 79,3333%.

PENDAHULUAN

Electronic Marketplace adalah sebuah tempat terjadinya transaksi antara penjual dan pembeli secara virtual [1]. Dalam proses transaksi jual beli, pihak penjual biasanya mengirimkan produk yang ingin dijual dan ditambahkan detail harga dan spesifikasi produk agar pihak pembeli dapat melihat spesifikasi produk yang sesuai dengan yang diinginkan seperti harga yang terjangkau tetapi kualitas produk yang baik. Dengan adanya e-marketplace dapat memungkinkan kegiatan jual beli terjadi tidak hanya di satu negara saja tetapi juga dapat terjadi di berbagai negara.

Pada e-marketplace terdapat biasanya terdapat fitur rating dan review produk sehingga sebelum melakukan transaksi pembelian, pihak pembeli melihat dulu kolom rating dan review dari produk tersebut untuk mengetahui kualitas dari produk tersebut. Akan tetapi terdapat kendala yaitu jumlah review yang sangat banyak sehingga untuk menilai produk itu memiliki kualitas yang baik menjadi sulit sehingga untuk melihat produk itu baik dilakukan dengan melihat rating yang diberikan oleh pengguna yang sudah membeli produk tersebut.

Rating produk yang diberikan dimulai dari bintang satu sampai dengan lima, dengan fitur ini pembeli dapat mengetahui kualitas dari produk itu berdasarkan banyaknya bintang yang diberikan.

Akan tetapi terdapat masalah seperti beberapa pembeli memberikan review negatif akan tetapi memberikan nilai *rating* sebesar lima bintang ataupun sebaliknya. Hal tersebut dapat menyebabkan fitur penilaian produk berdasarkan *rating* bintang menjadi kurang baik sehingga suatu produk tidak merepresentasikan nilai sesungguhnya.

Oleh karena itu, dibutuhkan suatu analisa sentimen untuk menyimpulkan ulasan terhadap produk secara tepat dan efisien (selain dari fitur *rating* tersebut). Pada penelitian ini dibuat analisis sentimen terhadap produk-produk yang terdapat pada situs *e-marketplace* berdasarkan ulasan produk dari pelanggan yang telah melakukan transaksi dengan produk tersebut dengan menggunakan metode TF-IDF dan KNN. Pada penelitian ini dibuat juga chrome ekstension yang menampilkan persentase hasil analisis sentimen dari produk yang digunakan. Produk yang digunakan merupakan produk permainan yang dijual pada situs *e-marketplace* shopee.

Beberapa penelitian telah dilakukan dibidang analisis sentimen pada *e-marketplace* seperti Rahayu yang melakukan analisis sentimen pada produk yang ada di *e-marketplace* shopee [2]. Data yang digunakan sebanyak 300 data latih dan 60 data uji yang terbagi menjadi dua kategori yaitu kategori positif dan kategori negatif. Tahapan selanjutnya dilakukan klasifikasi data uji menggunakan metode Naïve Bayes Classifier dan diperoleh akurasi sebesar 78,3% dan Fmeasure yang didapatkan sebesar 77,9 %.

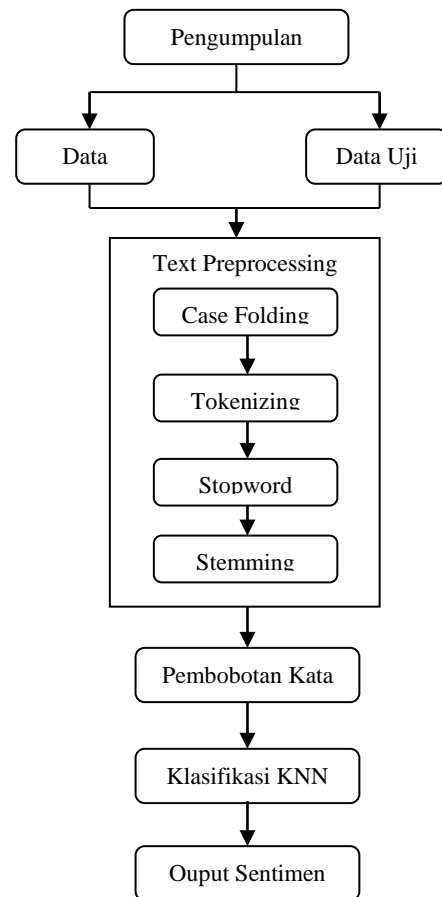
Penelitian selanjutnya, muktafin menggunakan metode TF-IDF dan KNN dengan pendekatan NLP untuk mengelompokkan review barang “hijab instan” ke dalam dua kategori (positif dan negatif). Dataset yang digunakan adalah 260 review dengan komposisi 208 data latih dan 52 data uji. Hasil klasifikasi menggunakan pendekatan NLP memperoleh akurasi sebesar 76,92%, presisi 80,00% dan recall 74,07% [3].

Pratmanto melakukan analisis sentimen pada aplikasi shopee yang ada pada google play. Data yang digunakan terdiri dari 140 data latih dan 60 data uji yang terbagi menjadi dua kategori yaitu positif dan negatif. Klasifikasi yang digunakan adalah klasifikasi dengan metode Naive Bayes dan diperoleh tingkat akurasi sebesar 96,667% [4].

Berdasarkan penelitian sebelumnya hanya dibentuk dua kategori sehingga pada penelitian ini dibuat analisis sentimen dengan tiga kategori yaitu positif, netral dan negatif.

METODE PENELITIAN

Pada bagian ini dibahas tahapan-tahapan penelitian dalam melakukan analisis sentimen produk permainan yang ada pada *e-marketplace* shopee yang ditunjukkan pada Gambar 1.



Gambar 1. Gambaran Umum Penelitian

Dalam penelitian ini, tahapan awal adalah mengumpulkan data untuk dijadikan sebagai data latih dan data uji. Tahapan selanjutnya adalah melakukan Text Preprocessing yaitu mengolah ulasan awal menjadi kata-kata yang dapat digunakan untuk menganalisis sentimen, setelah itu dilakukan pembobotan kata menggunakan perhitungan TF-IDF. Hasil perhitungan TF-IDF tersebut digunakan untuk mengklasifikasikan review yang ada pada data uji dengan metode K-Nearest Neighbor (KNN).

Pengumpulan Data

Pengumpulan data diperoleh dari ulasan produk permainan di website Shopee sebanyak 1000 data dan dibagi 700 data latih dan 300 data uji. Data tersebut terbagi menjadi tiga kategori yaitu positif, netral dan negatif. Tahapan selanjutnya adalah melakukan preprocessing pada data ulasan sehingga data tersebut dapat diolah pada tahapan berikutnya.

Text Preprocessing

Text preprocessing dilakukan untuk membersihkan data dari noise. Pada penelitian ini, tahapan preprocessing yang digunakan adalah case folding, tokenizing, stopwords dan stemming. Pada tahap case folding ulasan yang mempunyai huruf besar diubah menjadi huruf kecil, setelah itu dilakukan tahap tokenizing yaitu proses memisahkan kalimat menjadi kata-kata tunggal [5], [6].

Tahapan selanjutnya adalah melakukan filtering dengan melakukan stopwords removal yaitu setiap kata hasil tokenizing diperiksa, jika terdapat kata sambung, kata depan yang tidak ada hubungannya dalam analisis sentimen maka kata tersebut akan dihilangkan [7], [8]. Tahapan terakhir pada preprocessing adalah melakukan stemming yaitu kumpulan kata yang sudah dilakukan

proses stopword diubah menjadi bentuk kata dasar [9], [10]. Setelah tahapan preprocessing selesai dilakukan, selanjutnya dilakukan pembobotan kata.

Pembobotan Kata

Pembobotan kata merupakan proses pemberian nilai pada setiap kata yang telah melewati proses preprocessing. Pada penelitian ini pembobotan dilakukan dengan menggunakan metode TF-IDF yang ditunjukkan pada persamaan (1) dan (2). Pembobotan kata dilakukan untuk memberikan nilai kepada suatu kata untuk dijadikan input pada proses klasifikasi. Berikut ini tahapan pembobotan kata dengan menggunakan TF-IDF [11], [12].

- Menghitung jumlah term frequency (TF) tiap kata
Pada tahap ini, kalimat yang telah dipisah menjadi kata akan diberi nilai. Setiap kata yang muncul akan diberi nilai 1.
- Menghitung jumlah document frequency (DF) tiap kata
Pada tahap ini, dilakukan perhitungan document frequency (DF) dengan cara menjumlahkan nilai TF pada tiap kata.
- Menghitung jumlah inverse document frequency (IDF)
Setelah melakukan perhitungan term frequency (TF) langkah selanjutnya yaitu melakukan perhitungan inverse document frequency (IDF) yang ditunjukkan pada persamaan (1).

$$IDF(w) = \log\left(\frac{N}{DF(w)}\right) \quad (1)$$

- Menghitung bobot (Weight)
Langkah selanjutnya pada tahap TF-IDF adalah perhitungan bobot pada tiap kata. Mendapatkan hasil bobot dilakukan perkalian nilai TF dengan IDF Yang ditunjukkan pada persamaan 2.

$$W_{ij} = tf_{ij} \times \log(D/df_j) \quad (2)$$

Hasil pembobotan kata tersebut akan digunakan sebagai input pada proses klasifikasi dengan menggunakan K-Nearest Neighbor (KNN).

Klasifikasi K-Nearest Neighbor

Pada tahap ini, proses klasifikasi dilakukan dengan algoritma K-Nearest Neighbor terhadap data komentar ulasan yang telah melalui proses perhitungan bobot tiap kata. K-Nearest Neighbor dilakukan dengan mencari kelompok k objek dalam data latih yang paling dekat (mirip) dengan objek pada data baru atau data uji [13]. Dekat atau jauhnya jarak data latih yang paling dekat dengan objek yang akan diklasifikasi dapat dihitung dengan menggunakan metode cosine similarity yang ditunjukkan pada persamaan (3) [14], [15].

$$\cos(\theta_{QD}) = \frac{\sum_{i=1}^n Q_i D_i}{\sqrt{\sum_{i=1}^n (Q_i)^2} \sqrt{\sum_{i=1}^n (D_i)^2}} \quad (3)$$

Langkah-langkah dalam perhitungan Cosine Similarity yang harus dilakukan adalah sebagai berikut:

1. Menghitung perkalian skalar
2. Menghitung panjang vektor
3. Menghitung similarity

HASIL DAN PEMBAHASAN

Berdasarkan tahapan penelitian yang telah dilakukan, data yang dikumpulkan merupakan data review produk permainan yang ada pada situs belanja online shopee sebanyak 1000 data. Data tersebut dibagi

menjadi dua jenis yaitu 700 data menjadi data latih dan 300 data menjadi data uji. Contoh data yang telah dikumpulkan dapat dilihat pada Tabel 1.

Tabel 1. Contoh Data Ulasan *Customer* Shopee

Kode	Komentar
D1	Paket pesanan sudah sampai dan sudah saya terima dalam kondisi baik. Paket sesuai dengan deskripsi, thanks untuk Seller dan juga Shopee.
D2	Barangnya bagus pengiriman cepet dan aman bubble wrap dobel2
D3	Barang tidak sesuai di deskripsi, seller di chat berkali tidak ada respon, not recommended seller
D4	Barang sudah di terima tapi belum di coba apa berfungsi baik
D5	Mantap barangnya, cuman bungkusannya kurang tebal
D6	Terimakasih, paket telah sampai lengkap dan selamat. Produk berfungsi normal

Pada tahap berikutnya, dilakukan teks preprocessing seperti case folding, tokenizing, stopword dan stemming.

Case folding

Pada tahap ini, dilakukan proses dimana semua huruf diubah menjadi lowercase atau huruf kecil. Hasil case folding dapat dilihat pada Tabel 2.

Tabel 2. Hasil Case Folding

Kode	Komentar	Case Folding
D1	Paket pesanan sudah sampai dan sudah saya terima dalam kondisi baik. Paket sesuai dengan deskripsi, thanks untuk Seller dan juga Shopee.	paket pesanan sudah sampai dan sudah saya terima dalam kondisi baik. paket sesuai dengan deskripsi, thanks untuk seller dan juga shopee. 1
D2	Barangnya bagus pengiriman cepet dan aman bubble wrap dobel2	barangnya bagus pengiriman cepet dan aman bubble wrap dobel2 1
D3	Barang tidak sesuai di deskripsi, seller di chat berkali tidak ada respon, not recommended seller	barang tidak sesuai di deskripsi, seller di chat berkali tidak ada respon, not recommended seller - 1
D4	Barang sudah di terima tapi belum di coba apa berfungsi baik	barang sudah di terima tapi belum di coba apa berfungsi baik 0
D5	Mantap barangnya, cuman bungkusannya kurang tebal	mantap barangnya, cuman bungkusannya kurang tebal 0
D6	Terimakasih, paket telah sampai lengkap dan selamat. Produk berfungsi normal	terimakasih, paket telah sampai lengkap dan selamat. produk berfungsi normal 1

Tokenizing

Pada tahap ini, dilakukan proses pemisahan kalimat menjadi kata tunggal dan melakukan pengecekan kata dari karakter pertama sampai karakter terakhir. Hasil tokenizing dapat dilihat pada Tabel 3.

Tabel 3. Hasil Case Tokenizing

Kode	Case Folding	Tokenizing
D1	paket pesanan sudah sampai dan sudah saya terima dalam kondisi baik. paket sesuai dengan deskripsi, thanks untuk seller dan juga shopee.	[paket, pesanan, sudah, sampai, dan, sudah, saya, terima, dalam, kondisi, baik., paket, sesuai, dengan, deskripsi, thanks, untuk, seller, dan, juga, shopee.]
D2	barangnya bagus pengiriman cepet dan aman bubble wrap dobel2	[barangnya, bagus, pengiriman, cepet, dan, aman, bubble, wrap, dobel2]
D3	barang tidak sesuai di deskripsi, seller di chat berkali tidak ada respon, not recommended seller	[barang, tidak, sesuai, di, deskripsi., seller, di, chat, berkali, tidak, ada, respon, not, recommended, seller]
D4	barang sudah di terima tapi belum di coba apa berfungsi baik	[barang, sudah, di, terima, tapi, belum, di, coba, apa, berfungsi, baik]
D5	mantap barangnya, cuman bungkusannya kurang tebal	[mantap, barangnya., cuman, bungkusannya, kurang, tebal]
D6	terimakasih, paket telah sampai lengkap dan selamat. produk berfungsi normal	[terimakasih., paket, telah, sampai, lengkap, dan, selamat., produk, berfungsi, normal]

Stopword Removal

Pada tahap ini, kumpulan kata yang telah dilakukan proses tokenizing termasuk ke dalam tahap proses stopword removal yaitu setiap kata diperiksa, jika terdapat kata sambung atau kata yang tidak ada hubungannya dalam analisis sentimen, maka kata tersebut dihilangkan. Hasil stopword dapat dilihat pada Tabel 4.

Tabel 4. Hasil Stopword

Kode	Tokenizing	Stopword
D1	[paket, pesanan, sudah, sampai, dan, sudah, saya, terima, dalam, kondisi, baik., paket, sesuai, dengan, deskripsi, thanks, untuk, seller, dan, juga, shopee.]	[paket, pesanan, 'kondisi, baik, paket, sesuai, deskripsi, thanks, seller, shopee]
D2	[barangnya, bagus, pengiriman, cepet, dan, aman, bubble, wrap, dobel2]	[barangnya, bagus, pengiriman, cepet, aman, bubble, wrap, dobel2]
D3	[barang, tidak, sesuai, di, deskripsi., seller, di, chat, berkali, tidak, ada, respon, not, recommended, seller]	[barang, sesuai, deskripsi, seller, chat, berkali, respon, not, recommended, seller]
D4	[barang, sudah, di, terima, tapi, belum, di, coba, apa, berfungsi, baik]	[barang, terima, coba, berfungsi]
D5	[mantap, barangnya., cuman, bungkusannya, kurang, tebal]	[mantap, barangnya, cuman, bungkusannya, tebal]
D6	[terimakasih., paket, telah, sampai, lengkap, dan, selamat., produk, berfungsi, normal]	[terimakasih, paket, lengkap, selamat, produk, berfungsi, normal]

Stemming

Pada tahap ini, kumpulan kata yang sudah dilakukan proses stopword dilanjutkan kedalam tahap stemming yaitu pengubahan kata ke bentuk kata dasar atau penghapusan kata imbuhan. Hasil stemming dapat dilihat pada Tabel 5.

Tabel 5. Hasil Stemming

Kode	Stopword	Stemming
D1	[paket, pesanan, terima, 'kondisi, baik, paket, sesuai, deskripsi, thanks, seller, shopee]	[paket, pesan, terima, kondisi, baik, paket, sesuai, deskripsi, thanks, seller, shopee]
D2	[barangnya, bagus, pengiriman, cepet, aman, bubble, wrap, dobel2]	[barang, bagus, kirim, cepet, aman, bubble, wrap, dobel2]
D3	[barang, sesuai, deskripsi, seller, chat, berkali, respon, not, recommended, seller]	[barang, sesuai, deskripsi, seller, chat, kali, respon, not, recommended, seller]
D4	[barang, terima, coba, berfungsi]	[barang, terima, coba, fungsi]
D5	[mantap, barangnya, cuman, bungkusannya, tebal]	[mantap, barang, cuman, bungkus, tebal]
D6	[terimakasih, paket, lengkap, selamat, produk, berfungsi, normal]	[terimakasih, paket, lengkap, selamat, produk, fungsi, normal]

Pembobotan Kata

Pembobotan kata adalah suatu proses nilai terhadap setiap kata pada data ulasan yang telah melewati proses preprocessing. Pada penelitian ini digunakan metode TF-IDF seperti pada persamaan (1) dan (2) untuk proses pembobotan kata yang bertujuan untuk memberikan nilai kepada suatu kata yang dimana nilai dari kata tersebut menjadi input pada proses klasifikasi. Adapun contoh data latihan dan data uji dapat dilihat dari Tabel 6 dan Tabel 7.

Tabel 6. Contoh Data Latihan

Kode	Sebelum Preprocessing	Sesudah Preprocessing	Label
D1	Paket pesanan sudah sampai dan sudah saya terima dalam kondisi baik. Paket sesuai dengan deskripsi, thanks untuk Seller dan juga Shopee.	[paket, pesan, terima, kondisi, baik, paket, sesuai, deskripsi, thanks, seller, shopee]	Positif
D2	Barangnya bagus pengiriman cepet dan aman bubble wrap dobel2	[barang, bagus, kirim, cepet, aman, bubble, wrap, dobel2]	Positif
D3	Barang tidak sesuai di deskripsi, seller di chat berkali tidak ada respon, not recommended seller	[barang, sesuai, deskripsi, seller, chat, kali, respon, not, recommended, seller]	Negatif
D4	Barang sudah di terima tapi belum di coba apa berfungsi baik	[barang, terima, coba, fungsi]	Netral
D5	Mantap barangnya, cumanbungkusannya kurang tebal	[mantap, barang, cuman, bungkus, tebal]	Netral

Tabel 7. Contoh Data Uji

Kode	Sebelum Preprocessing	Sesudah Preprocessing	Label
D6	Terimakasih, paket telah sampai lengkap dan selamat. Produk berfungsi normal	[terimakasih, paket, lengkap, selamat, produk, fungsi, normal]	?

Tabel 6 merupakan contoh data latih yang sudah dilabelkan, sedangkan pada Tabel 7 merupakan data uji yang belum diketahui label sentimen. Adapun langkah-langkah dalam perhitungan bobot kata yang harus dilakukan adalah menghitung jumlah term frequency (TF) tiap kata.

Pada tahap ini, dilakukan perhitungan jumlah term frequency (TF), document frequency (DF) tiap kata dan menghitung jumlah inverse document frequency (IDF) menggunakan persamaan (1).

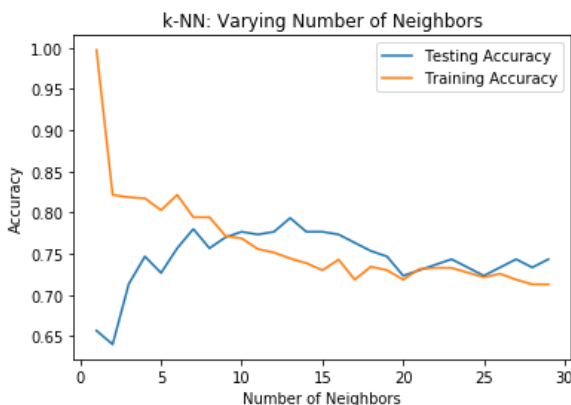
Hasil perhitungan TF-IDF dapat dilihat pada Tabel 8.

Tabel 8. Contoh Hasil Perhitungan TF-IDF

Kata	TF						DF	IDF
	D1	D2	D3	D4	D5	D6		
aman		0,7782					1	0,7782
bagus		0,7782					1	0,7782
baik	0,7782						1	0,7782
barang		0,1761	0,1761	0,1761	0,1761		4	0,1761
bubble		0,7782					1	0,7782
bungkus					0,7782		1	0,7782
cepat		0,7782					1	0,7782
chat			0,7782				1	0,7782
coba				0,7782			1	0,7782
cuman					0,7782		1	0,7782
deskripsi	0,4771		0,4771				2	0,4771
dobel2		0,7782					1	0,7782
fungsi				0,4771		0,4771	2	0,4771
kali			0,7782				1	0,7782
kirim		0,7782					1	0,7782
kondisi	0,7782						1	0,7782
lengkap						0,7782	1	0,7782
mantap					0,7782		1	0,7782
normal						0,7782	1	0,7782
not			0,7782				1	0,7782

Hasil Klasifikasi K-Nearest Neighbor

Tahapan selanjutnya adalah melakukan klasifikasi data uji berdasarkan data latih yang telah dilakukan preprocessing. Pada penelitian ini digunakan algoritma K-Nearest Neighbor (KNN) dengan memperhatikan banyaknya tetangga terdekat. Banyak tetangga terdekat yang digunakan adalah dari k = 1 sampai k = 29 yang ditunjukkan pada Gambar 2.



Gambar 2. Hasil Klasifikasi KNN Berdasarkan Jumlah Tetangga Terdekat

Berdasarkan pada Gambar 2, dapat dilihat bahwa nilai akurasi uji terbesar terjadi saat jumlah tetangga terdekat sebanyak k = 13 dengan tingkat akurasi sebesar 79,3333%. Tingkat akurasi tersebut diperoleh dari data pada Tabel 9.

Tabel 9. Hasil Prediksi Data Uji

		Prediksi		
		Negatif	Netral	Positif
Aktual	Negatif	87	14	8
	Netral	3	18	8
	Positif	7	22	133

Berdasarkan tabel 9, sentimen yang bernilai negatif pada keadaan sebenarnya diprediksi negatif sebanyak 87, diprediksi netral sebanyak 14 dan diprediksi positif sebanyak 8. Sentimen yang bernilai netral pada keadaan sebenarnya diprediksi bernilai negatif sebesar 3, diprediksi netral sebesar 18 dan diprediksi positif sebanyak 8. Selanjutnya, sentimen positif pada keadaan sebenarnya diprediksi negatif sebanyak 7, diprediksi netral sebanyak 22 dan diprediksi positif sebanyak 133.

Perhitungan tingkat akurasi menggunakan rumus seperti pada persamaan (4) [16].

$$\text{Tingkat akurasi} = \frac{TP}{TP+FP} \times 100\% \quad (4)$$

TP (True Positive) adalah hasil prediksi sesuai dengan data aktual (sebenarnya). FP (False Positive) adalah hasil prediksi tidak sesuai dengan data aktual (sebenarnya)

Berdasarkan Tabel 9 maka nilai $TP = 87 + 18 + 133 = 238$, nilai $FP = 14 + 8 + 3 + 8 + 7 + 22 = 62$ sehingga dengan menggunakan persamaan (4) diperoleh tingkat akurasi:

$$\begin{aligned} \text{Tingkat akurasi} &= \frac{238}{238+62} \times 100\% \\ &= \frac{238}{300} \times 100\% \\ &= 79,3333\% \end{aligned}$$

KESIMPULAN

Dalam melakukan belanja online pihak pembeli biasanya melihat terlebih dahulu ulasan dan rating yang diberikan dari pembeli sebelumnya agar dapat mengetahui kualitas dari produk tersebut. Akan tetapi sering terjadi kesalahan dalam pemberian rating seperti pembeli memberikan komentar negatif tetapi memberikan rating sebesar lima bintang ataupun sebaliknya. Oleh karena itu dibutuhkan metode lain agar pembeli dapat mengetahui produk tersebut memiliki kualitas yang bagus. Pada penelitian ini dilakukan analisis sentimen menggunakan metode TF-IDF dan K Nearest Neighbor (KNN) berdasarkan ulasan-ulasan dari pembeli untuk menyimpulkan ulasan terhadap produk secara tepat dan efisien. Data yang dikumpulkan adalah 1000 ulasan produk permainan pada situs belanja online shopee yang dibagi menjadi dua yaitu 700 data latih dan 300 data uji. Berdasarkan hasil penelitian diperoleh tingkat akurasi sebesar 79,3333%.

Untuk penelitian selanjutnya data yang digunakan diperbanyak tidak hanya produk permainan saja tetapi juga produk lainnya.

REFERENCES

- [1] Y. Kurniawan, S. E. Hiererra, H. Nicholas, J. Setiawan, and Frendy, "The Effect Of Indonesian E-marketplace Reputation On Customer Trust In Online Shopping," *JATIT*, vol. 98, no. 03, pp. 505–516, 2020.
- [2] S. Rahayu, Kusri, and H. Sismoro, "Sentimen Analisis Review Pengguna Marketplace Online Menggunakan Naive Bayes Classifier," *J. Inform. Interaktif*, vol. 3, no. 3, pp. 181–186, 2018.
- [3] E. H. Muktafin, Kusri, 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, 2020, doi: 10.30864/eksplora.v10i1.390.
- [4] D. Prاتمanto, R. Rousyati, F. F. Wati, A. E. Widodo, S. Suleman, and R. Wijianto, "App Review Sentiment Analysis Shopee Application In Google Play Store Using Naive Bayes Algorithm," *J. Physic Conf. Ser.*, vol. 1641, pp. 1–7, 2020, doi: 10.1088/1742-6596/1641/1/012043.
- [5] D. A. Muthia, "Analisis Sentimen Pada Review Restoran Dengan Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes," *J. Ilmu Pengetah. Dan Teknol. Komput.*, vol. 2, no. 2, pp. 39–45, 2017.
- [6] C. Fiami, H. Maharani, and R. Pratama, "Sentiment Analysis System for Indonesia Online Retail Shop Review Using Hierarchy Naive Bayes Technique," in *International Conference on Information and Communication Technologies (ICoICT)*, 2016, pp. 212–217.
- [7] A. Indriani, "Klasifikasi Data Forum dengan Menggunakan Metode Naive Bayes Classifier," in *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, 2014, pp. 5–10.
- [8] F. V. Sari and A. Wibowo, "Analisis Sentimen Pelanggan Toko Online JD.id Menggunakan Metode Naive Bayes Classifier Berbasis Konversi Ikon," *SIMETRIS*, vol. 10, no. 2, pp. 681–686, 2019.
- [9] Normah, "Naive Bayes Algorithm For Sentiment Analysis Windows Phone Store Application Reviews," *J. Publ. Informatics Eng. Res.*, vol. 3, no. 2, pp. 13–19, 2019.
- [10] F. Nurhuda, S. W. Sihwi, and A. Doewes, "Analisis Sentimen Masyarakat Terhadap Calon Presiden Indonesia 2014 Berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier," *J. ITSMART*, vol. 2, no. 2, pp. 35–42, 2013.
- [11] B. Herwijayanti, D. E. Ratnawati, and L. Muflikhah, "Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity," *Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 1, pp. 306–312, 2018.
- [12] R. Melita, V. Amrizal, H. B. Suseno, and T. Dirjam, "Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim)," *J. Tek. Inform.*, vol. 11, no. 2, pp. 149–164, 2018, doi: 10.15408/jti.v11i2.8623.
- [13] Murni, R. Kosasih, A. Fahrurrozi, T. Handhika, I. Sari, and D. P. Lestari, "Travel Time Estimation for Destination In Bali Using kNN-Regression Method with Tensorflow Travel Time Estimation for Destination In Bali Using kNN- Regression Method with Tensorflow," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 854, no. 012061, pp. 1–7, 2020, doi: 10.1088/1757-899X/854/1/012061.
- [14] M. M. Sya'bani and R. Umilasari, "Penerapan Metode Cosine Similarity dan Pembobotan TF / IDF pada Sistem Klasifikasi Sinopsis Buku di Perpustakaan Kejaksaan Negeri Jember," *J. Sist. Teknol. Indones.*, vol. 3, no. 1, pp. 31–42, 2018.
- [15] R. T. Wahyuni, D. Prastiyanto, and E. Suprptono, "Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi," *J. Tek. Elektro*, vol. 9, no. 1, pp. 18–23, 2017, doi: 10.15294/jte.v9i1.10955.
- [16] R. Kosasih, "Kombinasi Metode Isomap dan KNN Pada Image Processing Untuk Pengenalan Wajah," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 2, pp. 166–170, 2020.