

BAB 2

LANDASAN TEORI

2.1. Text Mining

Text adalah suatu hal umum yang di dalamnya terjadi pertukaran informasi. Informasi yang diambil harus menjadi data yang bermanfaat dan berguna. Oleh karena itu hal tersebut merupakan syarat umum data dan teks *mining*. *Text mining* adalah kegiatan mengidentifikasi teks untuk tujuan tertentu, sehingga informasi yang didapatkan menjadi berguna [5]. *Text mining* biasanya dimanfaatkan untuk pengkajian informasi, menentukan keputusan, dan tugas-tugas yang terkait dengan data teks dalam jumlah besar seperti manajemen informasi [7]. Pada awalnya, perolehan informasi (*information retrieval*), komputasi linguistik, data mining, statistik dan pembelajaran mesin (*machine learning*). Teks *mining* adalah bidang interdisiplin yang berpacu pada hal diatas. *Text mining* biasanya mencakup klasifikasi teks atau informasi, pengkategorian teks, pengembangan dan perumusan taksonomi umum, serta ekstraksi konsep atau entitas. *Text mining* mampu mengkategorikan dokumen dari kata-kata yang ada di dalamnya, menganalisa dokumen, serta menetapkan kesesuaian pada setiap dokumen untuk menelaah dengan cara apa mereka berhubungan antar setiap variabel [8].

2.2. Analisis Sentimen

Analisis sentimen adalah sebuah upaya untuk mengidentifikasi atau menganalisis sebuah pendapat seseorang yang memberikan penilaian terhadap suatu fenomena atau topik tertentu yang masuk kedalam kategori netral, negatif atau positif. Tidak seperti informasi aktual, sentimen dan opini mempunyai sifat yang memperlihatkan sebetulnya hal tersebut bersifat subjektif. Pengecekan gabungan opini dari berbagai aspek dibutuhkan karena hal tersebut merupakan penilaian subjektif yang awalnya lebih dari satu orang sehingga dibutuhkan sebuah ringkasan untuk mewakili suatu pendapat [3].

2.2.1. Analisis Sentimen Berbasis Aspek

Analisis sentimen yang dapat disebut juga *opinion mining* adalah wujud dari penerapan dibidang pemrosesan bahasa alami, linguistik komputasional, dan analisis teks yang dipakai untuk mengenali dan mengelompokan opini yang bersifat subjektif di dalam sumber. Meskipun pengelompokan pada level dokumen dan level kalimat sudah sangat bermanfaat, hasil dari analisis sentimen tidak begitu menghasilkan rincian yang diperlukan untuk dimanfaatkan pada kasus berbeda. Saat sebuah kalimat berisi pendapat dengan polaritas berbeda, hal tersebut dapat diartikan dengan analisis sentimen berbasis aspek atau ABSA. Berdasarkan SemEval 2014, tugas untuk ABSA terbagi menjadi 4 tugas, yaitu *aspect term extraction*, *aspect term polarity*, *aspect category detection*, dan *aspect category polarity* [9].

1. *Aspect term extraction* dan *aspect term polarity*

Pada tugas bagian ini, akan dilakukan pengenalan aspek yang berada pada kalimat yang diberikan.

2. *Aspect category detection*

Pada tugas bagian ini, akan dilakukan penggolongan pada label yang sudah ditetapkan.

3. *Aspect category polarity*

Pada tugas bagian ini, sudah terdapat klasifikasi aspek pada setiap ulasan.

2.3. Preprocessing

Tahap *preprocessing* dilakukan agar dataset lebih mudah digunakan untuk proses klasifikasi. Selain itu, tahapan *preprocessing* membuat dataset lebih seragam dan menghilangkan *noise* pada dataset. Tahapan dalam *preprocessing* adalah sebagai berikut:

2.3.1. Case Folding

Case folding adalah tahap merubah jika ada huruf kapital atau besar di dalam dataset jadi huruf kecil [10].

2.3.2. Cleaning

Cleaning adalah tahap membersihkan data seperti menghilangkan angka, tanda baca dan simbol dan emotikon pada dataset [2].

2.3.3. Tokenization

Tokenization adalah tahap mengubah kalimat menjadi token token pada dataset [2].

2.3.4. Word Normalization

Word Normalization adalah tahap membenarkan kata singkatan atau salah eja menjadi kata yang benar. Proses normalization ini berfungsi untuk merubah dimensi kata yang artinya sama tetapi memiliki ejaan yang salah [11].

2.3.5. Stopword Removal

Pada tahap ini, dilakukan pembuangan kata-kata yang dianggap tidak penting seperti: 'di', 'ke', 'dari', 'yang', 'dan', 'atau', 'ini', dan lainnya [12].

2.4. N-Gram

N-Gram adalah suatu potongan n kata yang didapatkan dari suatu *string*. Metode *N-Gram* ini biasanya diaplikasikan untuk pembangkitan kata atau karakter [13]. *N-gram* memiliki kelebihan dalam efisiensi dan efektifitas pengkategorian teks dan dapat mengklasifikasikan dengan akurasi yang tinggi [14]. Pemodelan *N-gram* adalah pendekatan identifikasi dan analisis fitur yang populer digunakan dalam pemodelan bahasa dan bidang pemrosesan bahasa alami. Model *N-gram* yang paling banyak digunakan dalam kategorisasi teks adalah *n-gram* berbasis *Word* dan *Characters*. Model *N-gram* dalam analisis sentimen membantu menganalisis sentimen teks atau dokumen. Karakteristik pada *N-gram* sebagai berikut [15]:

1. Walaupun terdapat kesalahan tekstual, *N-Gram* masih dapat berfungsi dengan baik.
2. Membutuhkan penyimpanan yang sederhana dan berjalan dengan efisien.
3. Waktu proses yang dibutuhkan relatif cepat.

Jenis *N-Gram* berdasarkan dari jumlah potongan gram *substring* yaitu *uni-gram*, *bi-gram*, *trigram*, *quad-gram* dan seterusnya sesuai dengan jumlah n dalam *n-gram* [15]. *N-Gram* yang umum digunakan adalah *Unigram*, *Bigram*, dan *Trigram*. *Unigram* adalah token yang terdiri dari satu kata, *Bigram* merupakan token yang terdiri dari dua kata, dan *Trigram* adalah *token* yang terdiri dari tiga kata. Contoh dari proses pemenggalan kalimat menggunakan *N-Gram Word* dengan jumlah pemotongan pada gram, jika hasil inputan “produknya tidak cocok dengan kulit saya” maka:

1. *Unigram*

{“produknya”, “tidak”, “cocok”, “dengan”, “kulit”, “saya”}

2. *Bigram*

{“produknya_tidak”, “tidak_cocok”, “cocok_dengan”, “kulit_saya”}

3. *Trigram*

{“produknya_tidak_cocok”, “tidak_cocok_dengan”, “cocok_dengan_kulit”, “dengan_kulit_saya”}

2.5. Pembobotan TF-IDF

Salah satu metode ekstraksi fitur yang paling populer untuk kasus klasifikasi teks adalah TF-IDF. Dalam analisis dokumen, dapat dilakukan penghitungan frekuensi kemunculan suatu kata dalam dokumen tertentu, yang disebut Term Frequency (TF), serta kemungkinan suatu kata muncul dalam berbagai dokumen, yang disebut Inverse Document Frequency (IDF). Dengan cara ini, dapat diukur tingkat kepentingan suatu kata dalam dokumen dan menentukan kata-kata yang paling relevan dalam memahami teks tersebut[16]. Metode ini digunakan untuk memberi bobot nilai berdasarkan tingkat kepentingan kata terhadap dokumen atau kategori dalam suatu kumpulan dokumen. Hasil dari TF-IDF adalah berupa matriks yang didapatkan dari hasil perkalian *Term Frequency* dengan *Inverse Document Frequency*. *Term Frequency* atau TF merupakan bobot nilai dari frekuensi term atau kata yang sering muncul pada dokumen. Sedangkan, *Inverse Document Frequency* atau IDF adalah jumlah dokumen terkait yang mengandung kata tertentu. Perhitungan TF-IDF dapat dilakukan menggunakan persamaan sebagai berikut [2]:

$$TF * IDF (d, t) = TF(d, t) * \log\left(\frac{N}{df(t)}\right) \quad (2.1)$$

Keterangan:

$TF * IDF (d, t)$: Pembobotan TF-IDF

$TF(d, t)$: Frekuensi munculnya term t pada dokumen d.

N : Jumlah dari semua kumpulan dokumen.

$df(t)$: Jumlah dari dokumen yang mengandung term t.

2.6. *K-Nearest Neighbor*

Metode *K-Nearest Neighbor* (K-NN) pertama kali di perkenalkan awal tahun 1960. Metode ini bekerja intensif ketika diberikan data training yang besar dan tidak populer sampai dengan tahun 1960. Ketika tekni komputasi sedang berkembang [5]. KNN merupakan algoritma klasifikasi *supervised learning*. KNN adalah salah satu algoritma sederhana dalam machine learning yang berbasis jarak. Perhitungan jarak dengan metode KNN dilakukan dengan cara menghitung jarak dari satu data yang berasal dari data test dengan seluruh data dari data train menggunakan *Euclidean Distance* [2] Terdapat beberapa tahapan untuk menentukan kelas atau label pada data, yaitu:

1. Menentukan nilai k
2. Menghitung jarak terdekat dengan persamaan *Euclidean Distance*
3. Menjumlahkan hasil perhitungan jarak
4. Mengurutkan hasil penjumlahan secara ascending
5. Memilih kategori tetangga terdekat sebanyak k
6. Memilih kelas berdasarkan frekuensi tertinggi terhadap k tetangga terdekat

Untuk menghitung jarak, dapat menggunakan persamaan *Euclidean Distance* :

$$deuclid = \sqrt{\sum_{i=1} |P_i - Q_i|^2} \quad (2.2)$$

Keterangan:

$deuclid$: jumlah fitur atau dimensi

P_i : fitur ke i pada data uji

Q_i : fitur ke i pada data latih

2.7. Confusion Matrix

Confusion Matrix berisi informasi mengenai performa dari suatu sistem klasifikasi yang dievaluasi menggunakan data atau metrik yang ada di dalam *Confusion matrix*. *Confusion Matrix* menganalisis seberapa baik klasifikasi yang telah dilakukan terhadap kelas aktual dan kelas hasil prediksi *matrix*. Hasil dari confusion matrix berupa nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Evaluasi yang akan digunakan yaitu *accuracy* [17].

Tabel 2.1. Confusion Matrix

Confusion Matrix		Prediksi		
		Positif	Non Aspek	Negatif
Aktual	Positif	TPP	PFNet	PFNeg
	Non Aspek	NetFP	TNetNet	NetFNeg
	Negatif	NegFP	NegFNeg	TNegNeg

Confusion Matrix menunjukkan tingkat akurasi dari proses klasifikasi yang telah dilakukan. Tingkat akurasi menunjukkan proporsi jumlah prediksi benar.

$$accuracy = \frac{TPP+TNetNet+TNegNeg}{Total} \quad (2.3)$$

2.8. Penelitian-Penelitian Terkait

Tabel 2.2. Penelitian-Penelitian Terkait

Review Literatur Pertama [18]	
Judul Artikel	Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek
Penulis	Jessica Widyadhana Iskandar, Yessica Nataliani
Judul Jurnal/Proceeding	Jurnal Resti (Rekayasa Sistem dan Teknologi Informasi)
Tahun Penerbitan	2021
Tujuan	Menganalisis sentimen terhadap sebuah produk gadget. Dengan menggunakan metode <i>Support Vector Machine</i> Naïve Bayes, SVM, dan k-NN.

Metode	Naïve Bayes, SVM, dan k-NN
Hasil Penelitian	Rata-rata accuracy SVM sebesar 96.43% dilihat dari empat aspek, yaitu aspek desain sebesar 94.40%, aspek harga sebesar 97.44%, aspek spesifikasi sebesar 96.22%, dan aspek citra merk sebesar 97.63%.
Kesimpulan	Rata-rata akurasi tertinggi dipegang oleh algoritma svm
Saran	melakukan pengujian dengan algoritma yang berbeda, dengan tidak hanya melakukan penelitian analisis sentimen komentar Youtube mengenai gadget, tetapi dengan topik yang berbeda. Selain melakukan penelitian dengan algoritma yang berbeda, dapat dilakukan juga analisis mendalam mengenai dataset yang akan dipakai. Selain itu, dapat juga dengan memperhatikan pelabelan setiap data yang akan dipakai. Pada tahap pre-processing perlu ditambahkan proses-proses lain untuk menghasilkan hasil klasifikasi dan prediksi yang lebih baik.
Review Literatur Kedua [3]	
Judul Artikel	PENERAPAN WORD N-GRAM UNTUK SENTIMENT ANALYSIS REVIEW MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (STUDI KASUS: APLIKASI SAMBARA)
Penulis	Fitriyani, Toni Arifin
Judul Jurnal/Proceeding	Jurnal Sistem Informasi
Tahun Penerbitan	2020
Tujuan	untuk mengetahui penerapan Word N-Gram, hasil nilai akurasi dengan menggunakan metode SVM, dan

	mengetahui seberapa besar pengaruh penerapan Word N-Gram terhadap nilai akurasi.
Metode Klasifikasi	Support Vector Machine, Word N-Gram
Hasil Penelitian	Hasil nilai akurasi tertinggi pada penelitian ini sebesar 89.00% dengan nilai AUC 0.944 (excellent classification) pada jumlah data 900, namun saat dilakukan penerapan Bi-gram dan Tri-gram menghasilkan penurunan akurasi. Nilai akurasi dengan kenaikan tertinggi yaitu pada penerapan Tri-gram dengan jumlah data 1.200. Kenaikan nilai akurasi sebesar 0.92% dibandingkan dengan Uni-Gram menjadi 88.59% dengan nilai AUC 0.954.
Kesimpulan	Hasil akurasi tertinggi pada penelitian ini sebesar 89.00% dengan nilai AUC 0.944 (excellent classification) pada jumlah data 900, namun saat dilakukan penerapan Bi-gram dan Tri-gram menghasilkan penurunan akurasi. Nilai akurasi dengan kenaikan tertinggi yaitu pada penerapan tri-gram dengan jumlah data 1.200. Kenaikan nilai akurasi sebesar 0.92% dibandingkan dengan Uni-Gram menjadi 88.59% dengan nilai AUC 0.954. Untuk itu dapat disimpulkan bahwa penerapan Word N-Gram akan meningkatkan nilai akurasi jika data yang digunakan dalam jumlah banyak.
Saran	Untuk penelitian selanjutnya disarankan membuat daftar stopword dari wordlist penerapan N-gram sebelum dilakukan proses klasifikasi. Hal ini mungkin bisa mengurangi dimensi yang akan membuat wordlist yang digunakan lebih unik dan memberikan nilai weight setiap katanya lebih besar.

	Melakukan pengembangan model penelitian yang lebih kompleks dengan menambahkan fitur – fitur yang ada, dan juga mencoba melakukan dengan metode yang lainnya untuk menemukan model yang terbaik.
Review Literatur Ketiga [19]	
Judul Artikel	PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DENGAN TF-IDF N-GRAM UNTUK TEXT CLASSIFICATION
Penulis	Nur Arifin, Ultach Enri, Nina Sulistiyowati
Judul Jurnal/Proceeding	STRING (Satuan Tulisan Riset dan Inovasi Teknologi)
Tahun Penerbitan	2021
Tujuan	Untuk mengklasifikasi artikel ilmiah ke dalam kategori sesuai dengan fokus dan ruang lingkup yang terdapat pada laman Syntax Jurnal Informatika secara otomatis dengan memanfaatkan proses text mining.
Metode	Support Vector Machine, N-Gram
Hasil Penelitian	Hasil penelitian setelah dilakukan pengujian terhadap model diukur dengan nilai Accuracy, Precision, Recall dan F-measure. Hasil terbaik adalah accuracy sebesar 70%, precision sebesar 75%, recall sebesar 69% dan f-measure sebesar 71% pada skenario perbandingan 90:10 dan kernel linear.
Kesimpulan	Penentuan jumlah minimal kata yang diproses dan fitur N-Gram berupa Unigram dan Bigram pada tahapan transformation dengan Term Frequency Inverse Document Frequency (TF-IDF), dinilai mampu meningkatkan nilai akurasi pada proses klasifikasi artikel ilmiah Syntax Jurnal Informatika. Nilai akurasi

	semula yaitu 30% dengan menentukan jumlah minimal kata yang diproses naik menjadi 60% dan dengan menerapkan fitur Unigram dan Bigram nilai akurasi berubah menjadi 70%.
Review Literatur Keempat [20] yg ini lg proses blm selesai di copy	
Judul Artikel	Integrasi N-gram, Information Gain, Particle Swarm Optimization di Naïve Bayes untuk Optimasi Sentimen Google Classroom
Penulis	Fajar Pramono, Didi Rosiyadi, Windu Gata
Judul Jurnal/Proceeding	Jurnal Resti(Rekayasa Sistem dan Teknologi Informasi)
Tahun Penerbitan	2019
Tujuan	Mencoba mencari nilai optimasi akurasi dari sentimen Google Classroom dengan mengintegrasikan N-gram, Information Gain, Particle Swarm Optimization (PSO) dan Klasifier Naïve Bayes yang belum pernah dilakukan para peneliti sebelumnya.
Metode	N-gram, Information Gain, Particle Swarm Optimization, Naïve Bayes
Hasil Penelitian	Dari percobaan yang dilakukan penggunaan N-gram dapat meningkatkan akurasi 6.7% dan AUC 4%, sedangkan menggunakan PSO mampu meningkat Akurasi 9.9% dan AUC 10.4%.
Kesimpulan	Penggunaan N-gram dan PSO sangat membantu dalam mengoptimalkan hasil Akurasi dan AUC, namun penambahan Information Gain dalam penelitian di sini tidak memberikan peningkatan yang berbeda,
Saran	Penggunaan parameter tambahan dan clasifier yang berbeda dapat memperbaiki peningkatan hasil akurasi dan AUC dengan seleksi fitur Information Gain.
Review Literatur Kelima [21]	

Judul Artikel	Aspect-Based Sentiment Analysis on Candidate Character Traits in Indonesian Presidential Election
Penulis	Rajalaxmi Hegde , Dr. Seema. S
Judul Jurnal/Proceeding	International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB17)
Tahun Penerbitan	2017
Tujuan	Untuk membandingkan kinerja iteratif yang diusulkan metode Random forest terhadap pembelajaran mesin lainnya seperti SVM, Baseline, dan K-Nearest Neighbor
Metode	K-Nearest Neighbor, Support vector machine (SVM) model
Hasil Penelitian	Berdasarkan hasil pengujian didapat hasil akurasi dari masing masing algoritma yaitu K-Nearest Neighbor 79%, Support vector machine 82%
Kesimpulan	Algoritma Support Vector Machine Memiliki performa yang lebih baik dari algoritma K-Nearest Neighbor karena memiliki akurasi yang lebih tinggi
Review Literatur Keenam [22]	
Judul Artikel	ANALISIS SENTIMENT PADA TWITTER DENGAN MENGGUNAKAN METODE K-Nearest Neighbor CLASSIFIER
Penulis	Sigit Suryono, Ema Utami, Emha Taufiq Luthfi
Judul Jurnal/Proceeding	PROSIDING
Tahun Penerbitan	2018
Tujuan	Untuk mengetahui polaritas dari sentimen positif, Non ataupun netral
Metode	K-Nearest Neighbor CLASSIFIER

Hasil Penelitian	Dari hasil uji coba sebanyak empat kali, didapatkan hasil tingkat akurasi pada percobaan pertama sebesar 62.98%, percobaan kedua sebesar 64.95%, percobaan ketiga sebesar 66.36%, dan percobaan keempat sebesar 66.79%. Dari hasil klasifikasi didapat tingkat persentase sentimen positif sebesar 28%, sentimen Non sebesar 20% dan sentiment netral sebesar 52%
Kesimpulan	Berdasarkan hasil persentase sentiment pengguna Twitter yang tidak memuji apa yang telah dilakukan oleh presiden Joko Widodo dan tidak juga menyalahkan apa yang telah dilakukan oleh presiden Joko Widodo. Dalam hal ini pengguna Twitter didominasi oleh sentiment neutral.
Review Literatur Ketujuh [23]	
Judul Artikel	Implementasi Algoritma K-Nearest Neighbor Terhadap Analisis Sentimen Opini Film Pada Twitter
Penulis	Fajar Ratnawati
Judul Jurnal/Proceeding	JURNAL INOVTEK POLBENG - SERI INFORMATIKA
Tahun Penerbitan	2018
Tujuan	Mengimplementasikan algoritma K-Nearest Neighbor Classifier terhadap analisis sentimen data opini film berbahasa Indonesia pada twitter
Metode	K-Nearest Neighbor
Hasil Penelitian	Berdasarkan hasil eksperimen, analisis sentimen yang dapat dilakukan oleh sistem dengan akurasi yang didapat adalah 90 % dengan rincian nilai precision 92%, recall 90% dan f-measure 90%.
Kesimpulan	Semakin banyak data training yang digunakan maka akan mempengaruhi kinerja dari sistem. Hasil akurasi

	akan semakin tinggi dan itu menandakan sistem berhasil melakukan klasifikasi dengan baik.
Review Literatur Kedelapan[15]	
Judul Artikel	Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan K-Nearest Neighbor Classifiers
Penulis	Taofik Krisdiyanto, Erry Maricha Oki Nurharyanto
Judul Jurnal/Proceeding	JURNAL CoreIT
Tahun Penerbitan	2021
Tujuan	Untuk mengetahui bagaimana sentimen masyarakat terhadap penerapan kebijakan PPKM di Indonesia melalui tweet dan komentar di platform media sosial Twitter.
Metode	K-Nearest Neighbor Classifiers
Hasil Penelitian	Hasil penelitian diketahui bahwa opini masyarakat mengenai kebijakan PPKM dihasilkan 99% termasuk kedalam klasifikasi polaritas positif dan 1% polaritas Non.
Kesimpulan	sentimen positif sebanyak 99% dan sentimen Non sebanyak 1% atau terlihat dari hasil klasifikasi emosi memiliki mayoritas “unknown” atau tidak diketahui yang sebagian besar kata tersebut memiliki polaritas positif
Review Literatur Kesembilan [12]	
Judul Artikel	Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode K-Nearest Neighbor
Penulis	Samsir, Ambiyar, Unung Verawardina, Firman Edi, Ronal Watrianthos

Judul Jurnal/Proceeding	JURNAL MEDIA INFORMATIKA BUDIDARMA
Tahun Penerbitan	2021
Tujuan	Menganalisis opini publik terhadap pembelajaran daring pada masa pandemi COVID-19 di Indonesia pada awal November 2020.
Metode	K-Nearest Neighbor
Hasil Penelitian	Temuan menunjukkan bahwa pembelajaran daring memiliki 30% sentimen positif, 69% sentimen Non, dan 1% netral pada periode tersebut.
Kesimpulan	Berdasarkan hasil ini pembelajaran daring belum maksimal diterapkan di Indonesia pada masa pandemi yang terlihat dari tingginya kekecewaan public pada awal November 2020.
Review Literatur Kesepuluh [24]	
Judul Artikel	Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode K-Nearest Neighbor
Penulis	Billy Gunawan, Helen Sasty Pratiwi, Enda Esyudha Pratama
Judul Jurnal/Proceeding	Jurnal Edukasi dan Penelitian Informatika
Tahun Penerbitan	2018
Tujuan	Untuk membantu perusahaan mengetahui umpan balik terhadap merek dagangnya dan masyarakat dalam menilai sebuah produk berdasarkan opini dan review yang ada.
Metode Klasifikasi	K-Nearest Neighbor
Hasil Penelitian	Hasil pengujian menunjukkan pada pengujian 3 kelas (Non, netral dan positif) hasil terbaik didapatkan pada 90% data latih dan 10% data uji dengan nilai akurasi

	77.78%, recall 93.33% dan precision 77.78% dan pada pengujian 5 kelas hasil terbaik didapatkan pada 90% data latih dan 10% data uji dengan nilai akurasi 59.33 %, recall 58.33 % dan precision 59.33 %
Kesimpulan	Sistem analisis sentimen pada ulasan produk online menggunakan metode K-Nearest Neighbor menghasilkan nilai akurasi terendah pada pengujian 5 kelas menggunakan dataset 80% latih dan 20% data uji sebesar 52.66%, sedangkan pada pengujian 3 kelas menggunakan dataset 90% data latih dan 10% data uji memiliki akurasi tertinggi sebesar 77.78%.