



UNIVERSITAS
MERCU BUANA

**PERBANDINGAN NAÏVE BAYES DAN KNN UNTUK ANALISIS
SENTIMEN SHOPEE BERBASIS ASPEK**

TUGAS AKHIR

**ANDRIANSYAH KAMAHIZA
41518110082**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA**

2022



UNIVERSITAS
MERCU BUANA

**PERBANDINGAN NAÏVE BAYES DAN KNN UNTUK ANALISIS
SENTIMEN SHOPEE BERBASIS ASPEK**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:

ANDRIANSYAH KAMAHIZA
41518110082

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA

2022

UNIVERSITAS
MERCU BUANA

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : 41518110082

Nama : Andriansyah Kamahiza

Judul Tugas Akhir : Perbandingan Naïve Bayes Dan KNN Untuk Analisis Sentimen Shopee Berbasis Aspek

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan di dalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.

Jakarta, 6 Agustus 2022



amahiza
(ANDRIANSYAH KAMAHIZA)

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Andriansyah Kamahiza
NIM : 41518110082
Judul Tugas Akhir : Perbandingan Naïve Bayes Dan Knn Untuk Analisis Sentimen Shopee Berbasis Aspek

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non Eksklusif** (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Non Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 6 Agustus 2022



Andriansyah Kamahiza

SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Andriansyah Kamahiza
NIM : 41518110082
Judul Tugas Akhir : Perbandingan Naïve Bayes Dan KNN Untuk Analisis Sentimen Shopee Berbasis Aspek

Menyatakan bahwa :

1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis		Status	
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi		Diajukan	
		Jurnal Nasional Terakreditasi	Sinta2		
		Jurnal Internasional Tidak Bereputasi		Diterima	
		Jurnal Internasional Bereputasi			
Disubmit/dipublikasikan di :	Nama Jurnal	: Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)			
	ISSN	: 2580-0760 (Online - Elektronik)			
	Link Jurnal	: http://jurnal.iaii.or.id/index.php/RESTI			
	Link File Jurnal Jika Sudah di Publish	:			

- Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit, revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
- Diminta untuk melampirkan scan KTP dan Surat Pernyataan (Lihat Lampiran Dokumen HKI), untuk kepentingan pendaftaran HKI apabila diperlukan

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 6 Agustus 2022


A F9070AKX022549117 niza
ANDRIANSYAH KAMAHIZA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518110082
Nama : Andriansyah Kamahiza
Judul Tugas Akhir : Perbandingan Naïve Bayes Dan KNN Untuk Analisis Sentimen Shopee Berbasis Aspek

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 6 Agustus 2022



(Rahmat Budiarto, Dr. Prof)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518110082
Nama : Andriansyah Kamahiza
Judul Tugas Akhir : Perbandingan Naïve Bayes Dan KNN Untuk Analisis Sentimen Shopee Berbasis Aspek

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 6 Agustus 2022



(Ir. Emil R. Kaburuan, Ph.D., IPM.)

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518110082
Nama : Andriansyah Kamahiza
Judul Tugas Akhir : Perbandingan Naïve Bayes Dan KNN Untuk Analisis Sentimen Shopee Berbasis Aspek

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 6 Agustus 2022



(Achmad Kodar, Drs. MT)

UNIVERSITAS
MERCU BUANA

LEMBAR LEMBAR PENGESAHAN

NIM : 41518110082
Nama : Andriansyah Kamahiza
Judul Tugas Akhir : Perbandingan Naïve Bayes Dan KNN Untuk Analisis Sentimen Shopee Berbasis Aspek

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 6 Agustus 2022

Menyetujui,



(Yaya Sudarya Triana, M.Kom., Ph.D.)
Dosen Pembimbing

Mengetahui,



(Wawan Guziawan, S.Kom, MT)
Koord. Tugas Akhir Teknik Informatika



(Ir. Emil R. Kaburuan, Ph.D., IPM.)
Ka. Prodi Teknik Informatika

MERCU BUANA

KATA PENGANTAR

Puji dan syukur kita panjatkan kehadirat Allah SWT, atas limpahan dan rahmat serta karunianya sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul “Perbandingan Naive Bayes dan KNN untuk analisis sentimen shopee berbasis aspek”. Penulis membuat Tugas Akhir ini karena merupakan salah satu syarat yang harus dipenuhi oleh mahasiswa tingkat akhir untuk menyelesaikan studi dan untuk mendapatkan gelar Sarjana Ilmu Komputer pada Program Studi Teknik Informatika Universitas Mercu Buana. Penulis menyadari bahwa tanpa bantuan dan bimbingan dari berbagai pihak tidak akan menyelesaikan tugas akhir ini dengan baik. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Kedua orang tua penulis , yang telah mendukung penuh penulis baik itu moril maupun materil selama 8 semester menuntut ilmu di perguruan tinggi.
2. Bapak Emil R. Kaburuan, Ph.D selaku Ketua Program Studi Teknik Informatika Universitas Mercu Buana.
3. Bapak Yaya Sudarya Triana, Ph.D selaku pembimbing proposal dan Tugas Akhir yang selalu sabar dalam membimbing dan mengajarkan penulis selama penyusunan proposal dan penyelesaian Tugas Akhir.
4. Bapak Herry Derajad Wijaya, S.Kom., MM selaku dosen pembimbing akademik penulis dari semester 1-6 yang selalu memberikan bimbingan akademik kepada penulis.
5. Bapak Muhammad Rifqi, S.Kom, M.Kom selaku dosen pembimbing akademik penulis saat ini yang juga memberikan bimbingan akademik dan memberikan arahan serta masukan kepada penulis.
6. Seluruh Jajaran Dosen dan Staf Program Studi Teknik Informatika Universitas Mercu Buana.
7. Keluarga besar penulis yang selalu mendukung dan mensupport serta memberikan semangat kepada penulis.

8. Teman sekaligus sahabat seperjuangan penulis selama menuntut ilmu di Universitas Mercubuana “6 Sekawan MercuBuana” , banyak hal yang kita lewati bersama dan banyak pula hal yang membuat kita bisa sampai dititik saat ini. Semoga kalian juga dapat menyelesaikan pendidikannya dengan baik dan semoga kalian selalu dilindungi oleh Tuhan YME.
9. Last but not least , terimakasih untuk diri sendiri , terimakasih selama 8 semester ini kamu telah berjuang sampai saat ini. Perjuanganmu masih panjang , masih banyak tujuan dan cita cita mu yang harus dicapai pada masa yang akan datang. Semangat Terus !!!

Akhir kata, penulis berharap semoga Tugas Akhir ini dapat bermanfaat dan menambah pengetahuan khususnya basgi penulis dan pembaca

Jakarta, 6 Agustus 2022

Penulis



UNIVERSITAS
MERCU BUANA

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL.....	ii
LEMBAR PERNYATAAN ORISINALITAS.....	iii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR... iv	
SURAT PERNYATAAN LUARAN TUGAS AKHIR	v
LEMBAR PERSETUJUAN PENGUJI.....	vi
LEMBAR PENGESAHAN.....	ix
ABSTRAK.....	x
ABSTRACT.....	xi
KATA PENGANTAR.....	xii
DAFTAR ISI.....	xiv
NASKAH JURNAL.....	1
KERTAS KERJA.....	8
RUMUSAN MASALAH.....	8
TUJUAN DAN MANFAAT.....	8
BATASAN MASALAH.....	9
BAB 1. LITERATUR REVIEW	10
BAB 2. ANALISIS DAN PERANCANGAN	21
BAB 3. SOURCE CODE	29
BAB 4. DATASET	43
BAB 5. TAHAPAN EKSPERIMEN.....	50
BAB 6. HASIL SEMUA EKSPERIMEN	60
DAFTAR PUSTAKA.....	68
LAMPIRAN DOKUMEN HAKI.....	69
LAMPIRAN KORESPONDENSI.....	71

NASKAH JURNAL

Terakreditasi SINTA Peringkat 2

Surat Keputusan Direktur Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Nomor: 158/E/KPT/2021
masa berlaku mulai Volume 5 Nomor 2 Tahun 2021 sampai Volume 10 Nomor 1 Tahun 2026

Terbit online pada laman web jurnal: <http://jurnal.iaii.or.id>



JURNAL RESTI

(Rekayasa Sistem dan Teknologi Informasi)

Vol. 6 No. x (2022) x - x

ISSN Media Elektronik: 2580-0760

Perbandingan Naïve Bayes dan *K-Nearest Neighbor* untuk Analisis Sentimen Shopee Berbasis Aspek

Comparison of Naïve Bayes and K-Nearest Neighbor for Aspect-Based Shopee Sentiment Analysis

Andrianysah Kamahiza¹, Yaya Sudarya Triana²

¹Teknik Informatika, Fakultas Ilmu Komputer, Universitas MercuBuana

²Sistem Informasi, Fakultas Ilmu Komputer, Universitas MercuBuana

1415181010109@student.mercubuana.ac.id, yaya.sudarya@mercubuana.ac.id*

Abstract

The Internet has grown so rapidly, it can make it easier for users to understand and retrieve information quickly, accurately, and covers more space and time. The rapid use of the Internet can be utilized by e-commerce companies to improve their performance and services. In Indonesia, there are many e-commerce that can be accessed easily through mobile platforms such as the Google Play store. There are Reviews on the Google Play Store to comment on apps or others that have been downloaded. This study aims to determine the sentiment of Shopee. Sentiment analysis is the process of using text analytics to obtain various data sources from the internet and various social media platforms. The goal is to get opinions from users on the platform. Aspect-based analysis is sentiment analysis that not only identifies how the sentiment is, but also determines where it is headed. In terms of the completeness of sentiment analysis, the aspect level can be said to be more comprehensive than the text or sentence level. This is because sentiment analysis at the sentence level only focuses on identifying positive and negative sentiments, without paying attention to the aspects they contain. To fully analyze it, it is necessary to know whether the sentiment is positive or negative in each aspect, so it is necessary to conduct a sentiment analysis based on that aspect. Based on the results of analysis and testing that has been carried out on comments on the Google Play Store on Shopee with a total of 5,000 comments, with aspects of application, delivery and payment. With the CRISP-DM model and comparing the Naïve Bayes (Naïve Bayes) and K-Nearest Neighbor (K-Nearest Neighbor) classification methods, it is proven that the Naïve Bayes (Naïve Bayes) classification method model shows the best results. The average Naïve Bayes Accuracy is 83.36% seen from three aspects, namely the application aspect of 79.30%, the delivery aspect of 82.14%, and the payment aspect of 86.96%

Keywords: Shopee, Sentiment Analysis, Naïve Bayes, K-Nearest Neighbor

Abstrak

Internet telah berkembang begitu pesat, dapat memudahkan pengguna untuk memahami dan mengambil informasi cepat, akurat, dan mencakup lebih banyak ruang dan waktu. Pesatnya penggunaan Internet dapat

Universitas Mercu Buana

dimanfaatkan oleh perusahaan-perusahaan e-commerce untuk dapat meningkatkan performa dan pelayanannya. Di Indonesia telah banyak e-commerce yang telah dapat diakses dengan mudah melalui platform mobile seperti pada Google Play store. Terdapat Ulasan pada *Google Play Store* untuk berkomentar tentang aplikasi atau lainnya yang telah diunduh. Penelitian ini bertujuan untuk mengetahui sentimen dari Shopee. Analisis Sentimen adalah proses penggunaan text analytics untuk mendapatkan berbagai sumber data dari internet dan berbagai platform media sosial. Tujuannya adalah untuk memperoleh opini dari pengguna yang terdapat pada platform tersebut. Analisis berbasis aspek adalah analisis sentimen yang tidak hanya mengidentifikasi bagaimana sentimen itu, tetapi juga menentukan ke mana arahnya. Dari segi kelengkapan analisis sentimen, level aspek dapat dikatakan lebih komprehensif daripada level teks atau kalimat. Hal ini dikarenakan analisis sentimen pada tataran kalimat hanya berfokus pada identifikasi sentimen Positif dan Negatif, tanpa memperhatikan aspek-aspek yang dikandungnya. Untuk menganalisis secara utuh perlu diketahui apakah sentimen Positif atau Negatif pada masing-masing aspek, sehingga perlu dilakukan analisis sentimen berdasarkan aspek tersebut. Berdasarkan hasil analisis dan pengujian yang telah dilakukan terhadap komentar pada *Google Play Store* terhadap Shopee dengan total 5,000 komentar, dengan aspek aplikasi, pengiriman dan pembayaran. Dengan model CRISP-DM dan membandingkan metode klasifikasi Naïve Bayes (*Naive Bayes*) dan *K-Nearest Neighbor* (*K-Nearest Neighbor*), terbukti bahwa model metode klasifikasi Naïve Bayes (*Naive Bayes*) menunjukkan hasil terbaik. Rata-rata *Accuracy Naive Bayes* sebesar 83.36% dilihat dari tiga aspek, yaitu aspek aplikasi sebesar 79.30%, aspek pengiriman sebesar 82.14%, dan aspek pembayaran sebesar 86.96%

Kata kunci: Shopee , Analisis Sentimen, Naïve Bayes, *K-Nearest Neighbor*

1. Pendahuluan

Dengan banyaknya pengguna internet tersebut merupakan peluang besar bagi perkembangan ekonomi di Indonesia. Ekonomi digital Indonesia terus mengalami perkembangan yang signifikan, dapat kita lihat dengan banyaknya jumlah e-commerce yang ada di Indonesia[1]. pemanfaatan internet yang cepat dapat digunakan oleh perusahaan e-commerce untuk meningkatkan kinerja dan layanan mereka. Di Indonesia telah banyak e-commerce yang dapat kita akses dengan mudah melalui platform mobile seperti pada Google Play Store. Salah satu cara untuk memanfaatkannya ialah dengan mengolah data komentar atau ulasan aplikasi e-commerce.

Teknik analisis sentimen merupakan proses dalam mengolah, memahami, dan mengekstrak data dalam bentuk teks terhadap suatu topik, kejadian ataupun individu untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini secara otomatis. Analisis sentimen memungkinkan untuk melacak produk, merek, maupun orang yang menentukan apakah hal tersebut dilihat sebagai suatu hal Positif dan Negatif[2].

Analisis sentimen dapat diaplikasikan pada perusahaan yang mengeluarkan suatu produk dan menyediakan layanan untuk menerima pendapat (feedback) dari konsumen untuk produk tersebut. Analisis sentimen diaplikasikan untuk mengelompokkan feedback Positif dan Negatif dari konsumen sehingga mempercepat dan mempermudah tugas perusahaan untuk meninjau kembali kekurangan

produk mereka[3]. Apabila ditemukan adanya sentimen Negatif, maka perusahaan dapat dengan cepat mengambil tindakan untuk menanggulangnya. Sebaliknya, jika sentimen Positif yang lebih dominan, perusahaan dapat meningkatkan produksi dan distribusi produk[2].

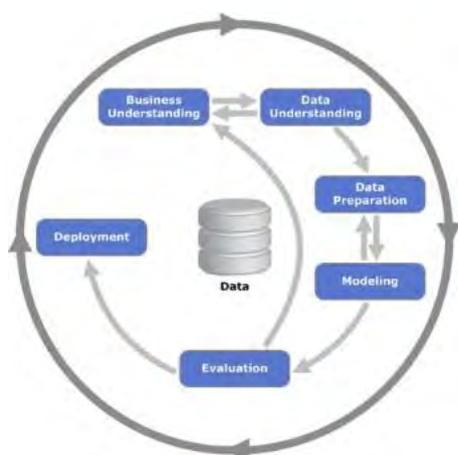
Penelitian dengan metode algoritma Naïve Bayes dan k nearest neighbor pernah dilakukan Josua Josen Alexander Limbong, Irwan Sembiring, kristoko dwi hartomo dengan judul Analisis Klasifikasi Sentimen Ulasan pada E-Commerce Shopee Berbasis Word Cloud dengan Metode Naive Bayes dan K-Nearest Neighbor menghasilkan metode Naive Bayes memperoleh hasil nilai accuracy 0,914, precision 0,915, recall 0,914 dan F1 score 0,916. Sedangkan metode KNN memperoleh nilai accuracy 0,928, precision 0,929, recall 0,928, dan F1 score 0,926.

Hal ini membuktikan bahwa dalam penelitian ini kinerja metode KNN lebih baik. Berdasarkan uraian diatas, tujuan dari penelitian ini adalah untuk melakukan analisis sentiment terhadap shopee berdasarkan ulasan komentar yang ada di play store dan membandingkan nilai akurasi kelima aplikasi tersebut dari model klasifikasi algoritma Naive Bayes dan k nearest neighbor untuk selanjutnya menentukan model klasifikasi terbaik berdasarkan nilai akurasi tersebut.

2. Metode Penelitian

2.1 Metode CRISP - DM

Metode penelitian ini menggunakan model CRISP-DM (Cross Industry Standard Process for Data Mining). CRISP-DM adalah metode yang menyediakan standar untuk data mining dan dapat diterapkan pada strategi pemecahan masalah umum. CRISP-DM juga merupakan standar untuk proses data mining sebagai strategi pemecahan masalah umum untuk bisnis atau penelitian [4]. Model CRISP-DM terdiri dari enam fase, yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment. Langkah-langkah CRISP-DM ditunjukkan pada Gambar 1. Penelitian ini dilakukan hanya pada tahap evaluasi dan tidak pada tahap implementasi, yaitu tahap implementasi suatu alat.



Gambar 1. Metode CRISP - DM

Model CRISP-DM diterapkan dalam penelitian ini, sehingga untuk tahapan penelitiannya disesuaikan menjadi seperti yang ditunjukkan pada Gambar 1. Tahapan penelitian terdiri dari pengumpulan data komentar shopee, dimana dataset yang digunakan bersumber dari ulasan pada *Google Play Store* mengenai Shopee, melakukan pelabelan data, pre-processing, cross validation, dan tahap evaluasi

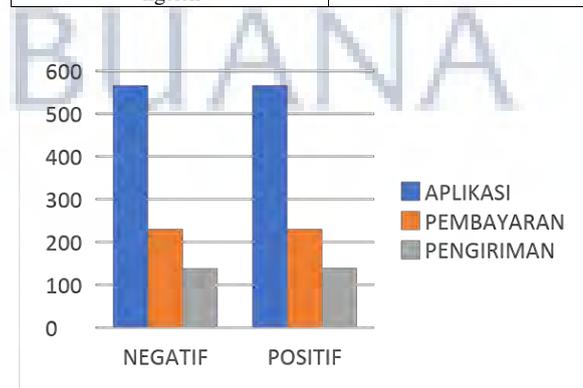
Tabel 1. Crawling Data

User Name	score	content
Leni Jpr	5	Dari dulu Bagus si menurut aku, setiap pembelian juga baik ² aja, bener gaada yang keliru lah ato apa. Pokoknya rekomended banget deh
Ika Pernalang	1	lemot, sering ngebut padahal sinyal kenceng+HP baik baik aja buat buka apk lain yang lebih berat aneh banget Shopee sering ngebug III HATEEEEE!!!!

Komentar yang tidak memiliki relevansi dengan penelitian tidak digunakan sebagai dataset. Setelah dilakukan seleksi, didapatkan sebanyak 1.870 komentar yang sesuai dengan penelitian. Dataset dilabeli secara manual, lalu dilakukan identifikasi aspek. Terdapat tiga aspek yang diteliti yaitu aplikasi, pengiriman, dan pembayaran. Aspek ditentukan berdasarkan hasil analisis komentar yang ada di kolom komentar google play store. Aspek aplikasi membahas mengenai update, kesan, dan pesan dalam penggunaan aplikasi. Aspek pengiriman membahas mengenai paket dan waktu dalam pengiriman paket. Aspek pembayaran membahas mengenai promo, voucher, dan pengembalian dana. Aspek yang didapatkan akan diklasifikasi menjadi duakelas, yaitu Positif ditandai sebagai 1 dan Negatif ditandai sebagai 0. Untuk komentar yang tidak mencakup lingkup penelitian akan ditandai 0. Contoh pelabelan suatu komentar terhadap masing-masing aspek dapat dilihat pada Tabel

Tabel 2. Pelabelan Data

content	polarity
Shopee dah ga asik buat seller. Udah ga bisa klik "Naikkan Produk" untuk dipromosin barang dagangan seller. JADI MALAS JUALAN DISHOPEE JADI MALAS BUKA APLIKASI SHOPEE Bulan depan aq uninstall aja deh.. Aplikasinya jadi kurang asikðŸ™™	Negatif
Banyakkk bangt vocernya aku suka belanja di sini dan ngak ribet, pengiriman juga selalu cepat dibanding toko online yang lain, dan barang selalu bagus sesuai expetasi ðŸ˜˜ðŸ˜˜ðŸ˜˜ðŸ˜˜ðŸ˜˜, cuman kalau pakai aplikasi ini sinyalnya kudu kuat kalau engak bakaln agak ngelek	Positif



Gambar 2. Grafik Pelabelan Data

2.1.3 Data Preparation

Langkah persiapan data dapat disebut sebagai langkah preprocessing. Langkah ini merupakan proses penyiapan clean data agar siap digunakan untuk penelitian. Beberapa pre-processing yang dilakukan antara lain Transform Case, yaitu mengubah huruf yang tidak beraturan dan tidak konsisten. Tokenisasi adalah proses menghilangkan tanda baca, simbol, karakter dan tanda baca yang dianggap tidak penting [5]. Normalisasi adalah proses perubahan kata yang tidak baku menjadi kata baku, menurut kosakata KBBI. Shortening adalah proses pengubahan kata dengan imbuhan menjadi kata dasar [6].

2.1.4 Modeling

Pada tahap ini dibuat model menggunakan klasifikasi untuk dataset komentar yang sudah diproses melalui tahap pre-processing. Penelitian ini menggunakan tools Google Colaboratory. Dua algoritma klasifikasi yang digunakan yaitu Naive Bayes dan K-Nearest Neighbor. Algoritma klasifikasi juga menggunakan bantuan SMOTE untuk mengatasi imbalance class[7].

2.1.5 Evaluation

Penelitian ini menggunakan cross-validation rating sebesar 1. Pada proses validasi terdapat dua sub proses untuk data latih, yaitu sub proses latih dan sub proses uji [6]. Setelah data training, hal ini akan mengarah pada evaluasi kinerja model klasifikasi dalam penelitian, yaitu Accuracy (Rumus 1), Precision (Rumus 2), Recall (Rumus 3) dan F-measure (Rumus 4).

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (2)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (3)$$

$$F - measure = \frac{(2 * TP * F1)}{(TP + F1)} \quad (4)$$

Ukuran evaluasi model klasifikasi terdapat TP (True Positive) yaitu hasil dari klasifikasi yang tepat, TN (True Negative) yaitu hasil dari klasifikasi yang tidak tepat, FP (False Positive) yaitu hasil dari klasifikasi yang tepat tetapi faktanya kurang tepat, FN (False Negative) yaitu hasil dari klasifikasi yang kurang tepat tetapi faktanya tepat.

1.1.1.1. 3. Hasil dan Pembahasan

3.1 Data Cleaning

3.1.1 Case folding

Case Folding merupakan proses dalam text preprocessing yang dilakukan untuk menyeragamkan karakter pada data. Proses Case Folding adalah proses mengubah seluruh huruf menjadi huruf kecil.

Tabel 3. Proses Case Folding

Sebelum	Sesudah
Shopee dah ga asik buat seller. Udah ga bisa klik "Naikkan Produk" untuk dipromosiin barang dagangan seller. JADI MALAS JUALAN DISHOPEE JADI MALAS BUKA APLIKASI SHOPEE Bulan depan aq uninstall aja deh.. Aplikasinya jadi kurang asik~!~!~!"	shopee dah ga asik buat seller. Udah ga bisa klik "naikkan produk" untuk dipromosiin barang dagangan seller. Jadi malas jualan dishopee jadi malas buka aplikasi shopee bulan depan aq uninstall aja deh.. aplikasinya jadi kurang asik~!~!~!"
UI AMPASSS, sering blank putih... gak kek mp yang lain normalw aja dan powerfull, Shopee bukannya benerin appnya malah promote terus... Perbaiki dulu lah sistem dan app kalian,... ~!~!~!"	ui ampasss, sering blank putih... gak kek mp yang lain normalw aja dan powerfull, shopee bukannya benerin appnya malah promote terus... perbaiki dulu lah sistem dan app kalian,... ~!~!~!"

3.1.2 Remove regex (cleansing) seperti tanda baca dan angka angka

angka dan tanda baca dalam kalimat tidak memiliki pengaruh pada text preprocessing. Tujuan tersebut agar hasil yang dicapai optimal. Menghapus tanda baca seperti [!"#\$%&'()*+,-./:;<=>?@[\] ^ _ ` { } ~]

Tabel 4. Proses Cleansing

Sebelum	Sesudah
shopee dah ga asik buat seller. udah ga bisa klik "naikkan produk" untuk dipromosiin barang dagangan seller. jadi malas jualan dishopee jadi malas buka aplikasi shopee bulan depan aq uninstall aja deh.. aplikasinya jadi kurang asik~!~!~!"	shopee dah ga asik buat seller udah ga bisa klik naikkan produk untuk dipromosiin barang dagangan seller jadi malas jualan dishopee jadi malas buka aplikasi shopee bulan depan aq uninstall aja deh aplikasinya jadi kurang asik~!~!~!"
ui ampasss, sering blank putih... gak kek mp yang lain normalw aja dan powerfull, shopee bukannya benerin appnya malah promote terus... perbaiki dulu lah sistem dan app kalian,... ~!~!~!"	ui ampasss sering blank putih gak kek mp yang lain normalw aja dan powerfull shopee bukannya benerin appnya malah promote terus perbaiki dulu lah sistem dan app kalian ~!~!~!"

3.1.3 Tokenize

Tokenisasi adalah proses untuk membagi teks yang dapat berupa kalimat, paragraf atau dokumen, menjadi token-token/bagian-bagian tertentu.

Tabel 5. Proses Tokenize

Sebelum	Sesudah
shopee dah ga asik buat seller udah ga bisa klik naikkan produk untuk dipromosiin barang dagangan seller jadi malas jualan dishopee jadi malas buka aplikasi shopee bulan depan aq uninstall aja deh aplikasinya jadi kurang asik	['shopee', 'dah', 'ga', 'asik', 'buat', 'seller', 'udah', 'ga', 'bisa', 'klik', 'naikkan', 'produk', 'untuk', 'dipromosiin', 'barang', 'dagangan', 'seller', 'jadi', 'malas', 'jualan', 'dishopee', 'jadi', 'malas', 'buka', 'aplikasi', 'shopee', 'bulan', 'depan', 'aq', 'uninstall', 'aja', 'deh', 'aplikasinya', 'jadi', 'kurang', 'asik']
ui ampasss sering blank putih gak kek mp yang lain normalw aja dan powerfull shopee bukannya benerin appnya malah promote terus perbaiki dulu lah sistem dan app kalian	['ui', 'ampasss', 'sering', 'blank', 'putih', 'gak', 'kek', 'mp', 'yang', 'lain', 'normalw', 'aja', 'dan', 'powerfull', 'shopee', 'bukannya', 'benerin', 'appnya', 'malah', 'promote', 'terus', 'perbaiki', 'dulu', 'lah', 'sistem', 'dan', 'app', 'kalian']

3.1.4 Stopwords

Stopwords adalah kata umum (*common words*) yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Contoh stopwords dalam bahasa Indonesia adalah “yang”, “dan”, “di”, “dari”, dll.

Tabel 6. Proses Stopwords

Sebelum	Sesudah
['shopee', 'dah', 'ga', 'asik', 'buat', 'seller', 'udah', 'ga', 'bisa', 'klik', 'naikkan', 'produk', 'untuk', 'dipromosiin', 'barang', 'dagangan', 'seller', 'jadi', 'malas', 'jualan', 'dishopee', 'jadi', 'malas', 'buka', 'aplikasi', 'shopee', 'bulan', 'depan', 'aq', 'uninstall', 'aja', 'deh', 'aplikasinya', 'jadi', 'kurang', 'asik']	shopee dah ga asik seller udah ga klik naik produk dipromosiin barang dagang seller malas jual dishopee malas buka aplikasi shopee aq uninstall aja deh aplikasi asik
['ui', 'ampasss', 'sering', 'blank', 'putih', 'gak', 'kek', 'mp', 'yang', 'lain', 'normalw', 'aja', 'dan', 'powerfull', 'shopee', 'bukannya', 'benerin', 'appnya', 'malah', 'promote', 'terus', 'perbaiki', 'dulu', 'lah', 'sistem', 'dan', 'app', 'kalian']	ui ampasss blank putih gak kek mp normalw aja powerfull shopee benerin appnya promote baik sistem app

3.1.5 Stemming

Stemming merupakan aplikasi pemotongan imbuhan pada kata berimbuhan yang dijalankan dengan algoritme tertentu.

Tabel 7. Proses Stemming

Sebelum	Sesudah
shopee dah ga asik seller udah ga klik naik produk dipromosiin barang dagang seller malas jual dishopee malas buka aplikasi shopee aq uninstall aja deh aplikasi asik	shopee dah ga asik seller udah ga klik naik produk dipromosiin barang dagang seller malas jual dishopee malas buka aplikasi shopee aq uninstall aja deh aplikasi asik
ui ampasss blank putih gak kek mp normalw aja powerfull shopee benerin	ui ampasss blank putih gak kek mp normalw aja powerfull shopee benerin

appnya promote baik sistem app	appnya promote baik sistem app
--------------------------------	--------------------------------

3.2 pembuatan model klasifikasi

3.2.1 Google Colaboratory

Tools yang sering digunakan pada pembuatan model klasifikasi ialah *Google Colaboratory*. *Google Colab* adalah *coding environment* bahasa pemrograman *Python* dengan format "notebook" (mirip dengan *Jupyter notebook*)[8].

3.2.2 Pemodelan K-Nearest Neighbor

K-Nearest Neighbors adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut[9].

```

# perform algoritma KNN
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
from sklearn.neighbors import KNeighborsClassifier

clf = KNeighborsClassifier(n_neighbors=7).fit(X_train, y_train)
predicted = clf.predict(X_test)
print('Confusion matrix:\n', confusion_matrix(y_test, predicted))
tn, fp, fn, tp = confusion_matrix(y_test, predicted).ravel()
print("TN: ", tn)
print("FP: ", fp)
print("FN: ", fn)
print("TP: ", tp)
print(classification_report(y_test, predicted, zero_division=0))
print("=====\n")
print("===== Hasil Klasifikasi Sentimen Analisis Shopee: =====")
print("Accuracy: ", accuracy_score(y_test, predicted))
print("Precision: ", precision_score(y_test, predicted, average="binary", pos_label="POSITIF"))
print("Recall: ", recall_score(y_test, predicted, average="binary", pos_label="POSITIF"))
print("F1_score: ", f1_score(y_test, predicted, average="binary", pos_label="POSITIF"))
print("error_rate: ", 1-accuracy_score(y_test, predicted))

confusion matrix:
[[158  31]
 [ 52 133]]
TN: 158
FP: 31
FN: 52
TP: 133

      precision    recall  f1-score   support

 NEGATIF      0.75      0.84      0.79      189
  POSITIF      0.81      0.72      0.76      185

 accuracy          0.78          0.78          0.78          374
 macro avg          0.78          0.78          0.78          374
weighted avg          0.78          0.78          0.78          374

=====

Hasil Klasifikasi Sentimen Analisis Shopee:
Accuracy: 0.7788748663101604
Precision: 0.8109756897568975
Recall: 0.7789189189189189
F1_score: 0.7621776584297995
error_rate: 0.22119513368983957

[11] from sklearn.model_selection import cross_val_score
from sklearn.neighbors import KNeighborsClassifier

[12] knn = KNeighborsClassifier(n_neighbors=3)
cross_val_score(knn, cos_sim, data[polarity], cv=10)

array([0.75478683, 0.7466821, 0.68449168, 0.75895629, 0.77885348,
       0.68449168, 0.79679144, 0.73262832, 0.74331551, 0.72843811])

[13] cross_val_score(knn, cos_sim, data[polarity], cv=10).mean()

0.7484922866136507
    
```

Gambar 3. Pemodelan K-Nearest Neighbor

3.2.3 Pemodelan Naive Bayes

Naive Bayes adalah metode yang cocok untuk klasifikasi biner dan multiclass. Metode yang juga dikenal sebagai Naive Bayes Classifier ini menerapkan teknik supervised klasifikasi objek di masa depan dengan menetapkan label kelas ke instance/catatan menggunakan probabilitas bersyarat[10].

```

from sklearn.model_selection import train_test_split
X=text_counts
y=data['polarity']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20, random_state=0)

[12] from sklearn.naive_bayes import ComplementNB
from sklearn.model_selection import GridSearchCV
from sklearn.model_selection import cross_val_score
from sklearn import metrics
from math import *
cnb = ComplementNB()
cnb.fit(X_train, y_train)
cross_cnb = cross_val_score(cnb, X, y, n_jobs = -1)
print("Cross validation score = ",cross_cnb)
print("Train accuracy = {:.2f}%".format(cnb.score(X_train,y_train)*100))
print("Test accuracy = {:.2f}%".format(cnb.score(X_test,y_test)*100))
train_acc_cnb=cnb.score(X_train,y_train)
test_acc_cnb=cnb.score(X_test,y_test)

Cross Validation score = [0.8287701 0.84204599 0.8368964 0.81598002 0.84966595]
Train accuracy =91.64%
Test accuracy =85.29%

[13] from sklearn.metrics import *
y_pred_cnb = cnb.predict(X_test)
from sklearn.metrics import confusion_matrix
print(confusion_matrix(y_test,y_pred_cnb))

[[[154  21]
 [  4 159]]]

[14] print(classification_report(y_test, y_pred_cnb))
roc_score_cnb=roc_auc_score(y_test, y_pred_cnb)
print("Area under the Curve = ",roc_score_cnb)

          precision    recall  f1-score   support

 0         0.83         0.89         0.86         185
 1         0.88         0.82         0.85         189

 accuracy         0.85         0.85         0.85         374
 macro avg         0.85         0.85         0.85         374
weighted avg         0.85         0.85         0.85         374

Area under the Curve = 0.8522961523961533

[15] from sklearn.metrics import *
print("F1 score = {:.2f}%".format(f1_score(y_test, y_pred_cnb, average="macro")*100))
f1_cnb=f1_score(y_test, y_pred_cnb, average="macro")
print("Precision score = {:.2f}%".format(precision_score(y_test, y_pred_cnb, average="macro")*100))
precision_cnb=precision_score(y_test, y_pred_cnb, average="macro")
print("Recall score = {:.2f}%".format(recall_score(y_test, y_pred_cnb, average="macro")*100))
recall_cnb=recall_score(y_test, y_pred_cnb, average="macro")

F1 score =85.29%
Precision score =85.45%
Recall score =85.33%
    
```

Gambar 4. Pemodelan Naive Bayes

3.2.4 Jumlah data masing masing aspek

Data akan melalui SMOTE untuk mengatasi imbalance class. Operator SMOTE Up Sampling berfungsi menyeimbangkan dataset menjadi seimbang[6]. Hasil dari proses operator SMOTE Up Sampling akan dilipatgandakan menggunakan operator Multiply. Proporsi data dapat dilihat pada Tabel .

Tabel 8. Jumlah data aspek

	Negatif	Positif	Grand Total
Aplikasi	566	566	1132
Pembayaran	230	230	460
Pengiriman	138	139	277
Grand Total	934	935	1869

3.2.5 Evaluasi model aspek

Setelah melewati proses validasi dengan algoritma klasifikasi Naive Bayes dan K-Nearest Neighbor, proses evaluasi akan menghasilkan tingkat Accuracy, Recall, Precision, dan Fmeasure dari masing-masing model. Berdasarkan Tabel , hasil Accuracy Naive Bayes lebih tinggi dibandingkan K-Nearest Neighbor. Aspek Aplikasi sebesar 79.30%, Aspek pengiriman 82.14%, dan aspek pembayaran sebesar 86.96%, sedangkan metode K-Nearest Neighbor dengan aspek aplikasi sebesar 71.37%, aspek pengiriman sebesar 71.37%, dan aspek pembayaran sebesar 63.04%

Tabel 9. Hasil Evaluasi Accuracy

Accuracy			
	aplikasi	pengiriman	pembayaran
K-NEAREST NEIGHBOR	71.37%	71.37%	63.04%
NAIVE BAYES	79.30%	82.14%	86.96%

Berdasarkan Tabel , hasil Precision Naive Bayes lebih tinggi dibandingkan K-Nearest Neighbor, Aspek Aplikasi sebesar 79.50%, Aspek pengiriman 81.25%, dan aspek pembayaran sebesar 86.95%, sedangkan metode K-Nearest Neighbor dengan aspek aplikasi sebesar 74.34%, aspek pengiriman sebesar 74.34%, dan aspek pembayaran sebesar 69.57%

Tabel 10. Hasil Evaluasi Precision

Precision			
	aplikasi	pengiriman	pembayaran
K-NEAREST NEIGHBOR	74.33%	74.33%	69.56%
NAIVE BAYES	79.50%	81.25%	86.95%

Berdasarkan Tabel , hasil Recall Naive Bayes lebih tinggi dibandingkan K-Nearest Neighbor, Aspek Aplikasi sebesar 78.74%, Aspek pengiriman 82.09%, dan aspek pembayaran sebesar 86.95%, sedangkan metode K-Nearest Neighbor dengan aspek aplikasi sebesar 70.00%, aspek pengiriman sebesar 70.00%, dan aspek pembayaran sebesar 61.54%.

Tabel 11. Hasil Evaluasi Recall

Recall			
	aplikasi	pengiriman	pembayaran
K-Nearest Neighbor	70.00%	70.00%	61.54%
Naive Bayes	78.74%	82.09%	86.95%

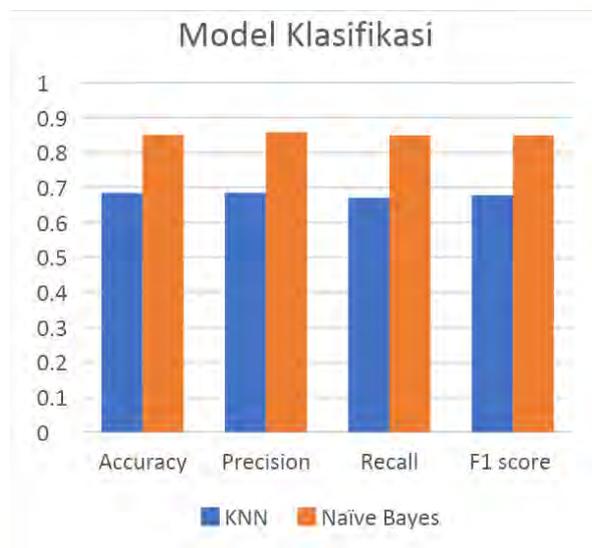
Berdasarkan Tabel , hasil F – measure Naive Bayes lebih tinggi dibandingkan K-Nearest Neighbor,

Aspek Aplikasi sebesar 78.93%, Aspek pengiriman sebesar 81.55%, dan aspek pembayaran sebesar 86.95%, sedangkan metode *K-Nearest Neighbor* dengan aspek aplikasi sebesar 72.10%, aspek pengiriman sebesar 72.10%, dan aspek pembayaran sebesar 61.31%.

Tabel 12. Hasil Evaluasi *F – measure*

<i>F – measure</i>			
	aplikasi	pengiriman	pembayaran
<i>K-NEAREST NEIGHBOR</i>	72.10%	72.10%	65.31%
<i>NAIVE BAYES</i>	78.93%	81.55%	86.95%

3.2.6 Perbandingan rata rata model klasifikasi



Gambar 5. Grafik Perbandingan Hasil Evaluasi Aspek

1.1.1.2. 4. Kesimpulan

Berdasarkan hasil analisis dan pengujian yang telah dilakukan terhadap ulasan shopee pada *Google Play Store*, lebih banyak pengguna yang memberikan opini Negatif dalam segala aspek dari aplikasi, pengiriman maupun pembayaran. Dengan membandingkan metode *Naive Bayes* dan *K-Nearest Neighbor* terbukti bahwa model *Naive Bayes* menunjukkan hasil terbaik. Rata-rata *Accuracy Naive Bayes* sebesar 84.26% dari tiga aspek, yaitu aspek aplikasi sebesar 79.30%, aspek pengiriman sebesar 82.14%, aspek spesifikasi sebesar 86.96%

Daftar Rujukan

- [1] I. R. Aysa, «Tantangan Transformasi Digital Bagi Kemajuan Perekonomian Indonesia», *J. At-Tamwil*

Kaji. Ekon. Syariah, 卷 3, 期 2, 页 140–153, 2021.

[2] D. Y. Praptiwi, « Analisis Sentimen Online Review Pengguna E-Commerce Menggunakan Metode Support Vector Machine Dan Maximum Entropy (Studi Kasus: Review Bukalapak pada Google Play) », 2018.

[3] F. M. Matulatuwa, E. Sedyono 和 A. Iriani, « Text mining dengan metode lexicon based untuk sentiment analysis pelayanan PT. Pos Indonesia melalui media sosial Twitter », *J. Masy. Inform. Indones.*, 卷 2, 期 3, 页 52–65, 2017.

[4] A. D. A. Putra 和 S. Juanita, « Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN », *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, 卷 8, 期 2, 页 636–646, 2021.

[5] D. Rustiana 和 N. Rahayu, « Analisis Sentimen Pasar Otomotif Mobil: Tweet Twitter Menggunakan Naive Bayes », *Simetris J. Tek. Mesin, Elektro Dan Ilmu Komput.*, 卷 8, 期 1, 页 113–120, 2017.

[6] J. W. Iskandar 和 Y. Nataliani, « Perbandingan Naive Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek », *J. RESTI (Rekayasa Sist. Dan Teknol. Informasi)*, 卷 5, 期 6, 页 1120–1126, 2021.

[7] A. N. Rais 和 A. Subekti, «Integrasi SMOTE Dan Ensemble AdaBoost Untuk Mengatasi Imbalance Class Pada Data Bank Direct Marketing », *J. Inform.*, 卷 6, 期 2, 页 278–285, 2019.

[8] R. T. Handayanto 和 H. Herlawati, «Machine Learning Berbasis Desktop dan Web dengan Metode Jaringan Syaraf Tiruan Untuk Sistem Pendukung Keputusan», *J. Komtika (Komputasi dan Inform.)*, 卷 4, 期 1, 页 15–26, 2020.

[9] A. M. Argina, « Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes », *Indones. J. Data Sci.*, 卷 1, 期 2, 页 29–33, 2020.

[10] R. S. Legowo, « Klasifikasi gerakan tangan SIBI (sistem isyarat bahasa indonesia) menggunakan leap motion dengan metode klasifikasi naive bayes ». Institut Teknologi Sepuluh Nopember, 2017.

KERTAS KERJA

Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal dengan judul di atas. Di dalam kertas kerja ini disajikan:

1. Literature Review
2. Analisis dan Perancangan
3. Source Code
4. Dataset yang digunakan
5. Tahapan Eksperimen
6. Hasil Eksperimen

RUMUSAN MASALAH

Berdasarkan uraian dari latar belakang, maka permasalahan dapat dirumuskan sebagai berikut:

1. Seberapa optimal metode Klasifikasi Naives Bayes dan K-Nearest Neighbor dalam analisis sentimen berbasis aspek?
2. Seberapa besar perbandingan metode Klasifikasi Naives Bayes dan K-Nearest Neighbor dalam analisis sentimen berbasis aspek?
3. Bagaimana Metode Klasifikasi *Naive Bayes dan K-Nearest Neighbor* dalam menganalisis sentimen berbasis aspek ?

TUJUAN DAN MANFAAT

A. Tujuan

Berdasarkan rumusan masalah diatas, maka tujuan pada penelitian ini adalah sebagai berikut :

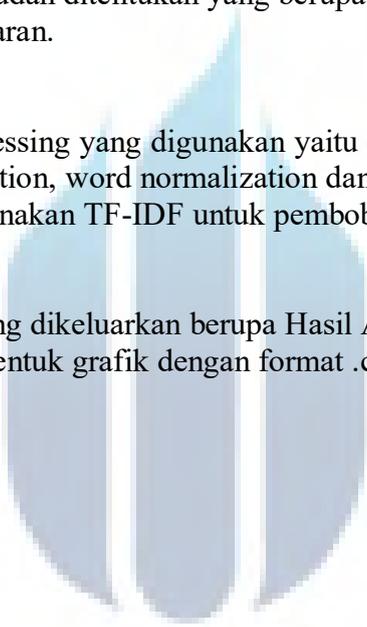
1. Mengetahui penerapan metode Klasifikasi Naives Bayes dan K-Nearest Neighbor dalam analisis sentimen berbasis aspek

Mengetahui perbandingan metode Klasifikasi Naives Bayes dan K-Nearest Neighbor dalam analisis sentimen

BATASAN MASALAH

Agar penelitian ini dapat dilakukan lebih fokus dan terarah, maka dibuat batasan masalah sebagai berikut:

1. Data Masukan
 - a. Format file dataset adalah .csv.
 - b. Data yang digunakan adalah dataset hasil scrapping data google play store
 - c. Hanya mengambil ulasan aplikasi Shopee pada Google Play Store khususnya kategori Ulasan Paling relevan
 - d. Teks yang diklasifikasikan dalam Bahasa Indonesia.
 - e. Aspek sudah ditentukan yang berupa aspek aplikasi, pengiriman, pembayaran.
2. Proses
 - a. Preprocessing yang digunakan yaitu casefolding, filtering, tokenization, word normalization dan stopword removal.
 - b. Menggunakan TF-IDF untuk pembobotan kata.
3. Data Keluaran
 - a. Data yang dikeluarkan berupa Hasil Akurasi yang akan disimpan dalam bentuk grafik dengan format .csv.



UNIVERSITAS
MERCU BUANA